DOCUMENT RESUME

ED 080 416 SO 006 039

AUTHOR	Semb, George, Ed.; And Others
TITLE	Behavior Analysis and Education.
INSTITUTION	Kansas Univ., Lawrence. Dept. of Human Development.
PUB DATE	72
NOTE	442p.
AVAILABLE FROM	Department of Human Development, University of Kansas, Lawrence, Kansas 66044 ($4.98)
EDRS PRICE	MF-$0.65 HC Not Available from EDRS.
DESCRIPTORS	*Behavioral Science Research; *Behavioral Sciences; Behavior Patterns; Case Studies (Education); Child Development; Conference Reports; Educational Objectives; *Educational Programs; Educational Strategies; Educational Theories; Elementary Grades; Higher Education; Intermediate Grades; Secondary Grades; State of the Art Reviews; Student Behavior; Teacher Education

ABSTRACT

This book presents part of the proceedings from the Third Annual Conference on Behavior Analysis in Education which was designed to provide Project Follow Through with the most current research products in the field, and to serve participating researchers by providing a medium for exchange of information regarding the state of the art and its most immediate challenges. The papers published are organized in seven sections. Part I includes a discussion of some of the current problems confronting applied behavior analysis and suggestions of some alternate techniques behavior analysis might use to have a greater impact on education. Part II consists of seven papers which address themselves to relatively new areas of research. Subsequent sections present a comprehensive behavior analysis program for handicapped children, description of the use of students as behavioral engineers, consideration of teacher training programs, current techniques and procedures. The concluding section emphasizes a behavioral approach to higher education. (Author/SHM)

ED 080416

behavior analysis & education

-- 1972

EDITED BY:

GEORGE SEMB

with:
DONALD R. GREEN
ROBERT P. HAWKINS
JACK MICHAEL
ELERY L. PHILLIPS
JAMES A. SHERMAN
HOWARD SLOANE
DON R. THOMAS

SP006039

FILMED FROM BEST AVAILABLE COPY

The project presented or reported herein was performed pursuant to a Grant from the U.S. Office of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the U.S. Office of Education, and no official endorsement by the U.S. Office of Education should be inferred.

ED 080416

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICROFICHE ONLY HAS BEEN GRANTED BY
George Semb
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER

Behavior Analysis & Education --1972

EDITED BY: **George Semb**

with: Donald R. Green
Robert P. Hawkins
Jack Michael
Elery L. Phillips
James A. Sherman
Howard Sloane
and Don R. Thomas

THE UNIVERSITY OF KANSAS
SUPPORT AND DEVELOPMENT CENTER
FOR FOLLOW THROUGH
DEPARTMENT OF HUMAN DEVELOPMENT
1972

This book presents part of the proceedings from the Third Annual Conference on Behavior Analysis in Education. It was sponsored by the Behavior Analysis Model for Follow Through, whose director is DON BUSHELL, JR., Department of Human Development, University of Kansas. The conference was supported in part by grants from the U.S. Office of Education as a component of PROJECT FOLLOW THROUGH. With FOLLOW THROUGH support, several communities throughout the country have implemented Behavior Analysis programs in elementary classrooms which serve low-income children. The responsibility for insuring academic success for thousands of children served by these projects is too great to be assumed solely by a single organization. It requires the best technical products of the entire professional community of Applied Behavior Analysis. This conference was designed to provide FOLLOW THROUGH with the most current research products in the field, and to serve participating researchers by providing a medium for the exchange on information regarding the "state of the art" and its most immediate challenges.

DEDICATION

To the thousands of *FOLLOW THROUGH* children who have benefited from the Behavior Analysis Program, their parents and their teachers, and the millions of others who may benefit in the future.

ACKNOWLEDGMENTS

One individual, more than any other, deserves special recognition for her outstanding services as typist, editor, English grammarian, proofreader, artist and organizer--*Cindy Stewart*. With almost a single-handed effort, Cindy has made this book possible. We frequently talk about errorless performance, but seldom do we encounter an individual whose dedication to a task produces it. Cindy Stewart is one of those rare people, and it was our fortune that she was responsible for the final preparation of this book.

The co-editors of this book deserve special credit, too, for their long hours of editorial review, committee meetings and the conference itself. We received over 120 manuscripts, each of which was reviewed by at least two of the co-editors. Nearly every manuscript accepted for presentation at the conference underwent at least one revision. Those which are presented here have been further edited and revised. The co-editors did an outstanding job to insure that the final product would be our best product.

We are also indebted to *Gene Ramp* and *Bill Hopkins*, authors of the 1971 proceedings, for their support and encouragement. Their experiences of a year ago enabled us to avoid many of the problems they encountered.

Finally, we would like to express our thanks to the many people who helped make this conference a success because of their many, generally unacknowledged contributions. These include those who helped transport visitors to and from the airport and campus, the many people around town who lent their rooms, yards, sofas and homes to students who attended the conference without financial support, members of the Follow Through staff who helped with innumerable tasks from audio-visual services to entertainment, the entire Department of Human Development, and most importantly the conference participants who critically but warmly helped us and each other.

FOREWORD

The Third Annual Conference on Behavior Analysis in Education met on the Lawrence campus of the University of Kansas on May 10, 11 and 12, 1972. This book is a compilation of several of the papers and discussions presented at the conference.

A little over a year ago, Gene Ramp and Bill Hopkins published *A NEW DIRECTION FOR EDUCATION: BEHAVIOR ANALYSIS 1971*. In their foreword, they outlined the role of applied behavior analysis in education. First, they stressed that the field deals with *behavior* - the observable things that school children, teachers, administrators and parents do. Second, they noted how the field has adopted an approach to education which emphasizes the design, organization and implementation of programs which produce better results than traditionally have been found. Finally, they presented several papers which demonstrated how procedures derived from the scientific community - experimentation and empirical analysis - served to improve educational practices and procedures.

The papers published in the present book expand upon those themes. If the field is to say that it has models that offer more and produce more for the educational community, then it must maintain its scientific base, and it must continue to attack new problems. The papers published here demonstrate not only how techniques and procedures can be analyzed systematically, but also how they can be used to solve practical problems in education.

Along another important dimension, Ramp and Hopkins also considered the issue of which behaviors ought to be valued most. They concluded that we have not yet found broadly acceptable solutions, but that we should actively continue to search for answers to the question. As Professor Baer noted in his 1971 address, critics no longer argue that behavior cannot be modified, but rather that it should not be modified. The critics are concerned that if you change a child's behavior, you may also change his inner essence. These are the same critics who argued just a few years ago that behavior modification would not work because behavior was simply a symptom of underlying causes and motives. Sometimes, it is reinforcing to point out inconsistencies in your opponent's position. It is difficult, however, to ignore the larger questions he raises, namely control and accountability within the profession. The panel discussion presented here addresses itself to those issues.

We have arbitrarily divided the conference proceedings into seven parts. The first, *Behavior Analysis: Past, Present and Future,* consists of three papers. Although it may seem strange to talk about the past in a field which is so young, Deborah Nedelman and Stephen Sulzbacher bring back pleasant memories of a boy named Dicky who was the subject of one of the classic studies of applied behavior analysis with children. Nedelman and Sulzbacher have compiled a ten year follow-up of Dicky and his progress through school which is a pleasure to read.

Some of the current problems confronting applied behavior analysis are also considered in Part 1. A panel of experts in the field discussed the question of standardized training programs and professional licensing. Although no consensus was reached, several alternatives were proposed. Finally in Part 1, Professor Becker suggests some alternate techniques behavior analysis might use to have a greater impact on education. These include: 1) broader public and political support for behavior analysis; 2) development of effective, large-scale educational programs with data to support them; and 3) a greater role in teacher training at the university level. Becker presents not only ideas, but also data from his Behavior Analysis Follow Through Program.

Part 2, *Current Topics in Behavior Analysis,* consists of seven papers which address themselves to relatively new areas of research. Topics include the effects of drugs on classroom behavior, creativity in preschoolers' art work, the development of freedom and responsibility in elementary school children, the development of phonetic sounds in preschool children, an analysis of open classroom teaching procedures and finally, the integration of normal and handicapped children in a behavior modification preschool.

Parts 3, 4, 5 and 6 all describe current programs and techniques in behavior analysis. Part 3 presents a comprehensive behavior analysis program for handicapped children. Part 4 includes four papers which describe the use of students as behavioral engineers. Part 5 considers teacher training programs, the effects of follow-up schedules on the maintenance of a teaching program and a description of a consulting teacher's success with a teacher training package. Part 6 is a mixture of current techniques and procedures which include a homeward educational program for a predelinquent juvenile, an auto-pedestrian safety program for elementary school children, the use of delayed consequences to control attendance and disruptive behaviors in a special education class, an analysis of three punishment techniques to control disruptive behavior, the use of reinforcement contingencies and teacher aides to alter subtraction performance, and finally, the effects of superimposing contracts on a token economy.

Part 7, *Behavior Analysis in Higher Education*, emphasizes a behavioral approach to higher education. The Born et al. study compares the study time investment of students in a Keller-style, personalized instruction course with that of students in a more traditional lecture course. Johnston and his colleagues provide a fine grain analysis of study behavior as it occurs in the natural environment with data collected from self-report inventories. The Miller and Weaver paper demonstrates how a multiple baseline design can be used to evaluate teaching effectiveness. Finally, Bostow and Blumenfeld report two experiments concerned with evaluating alternative procedures for awarding points to students on the basis of quiz performance.

TABLE OF CONTENTS

page

Foreword . . . . **vii**

PART 1: *BEHAVIOR ANALYSIS: PAST, PRESENT and FUTURE* **1**

- DICKY AT 13 YEARS OF AGE: A LONG TERM SUCCESS FOLLOWING EARLY APPLICATION OF OPERANT CONDITIONING PROCEDURES **3**

  Deborah Nedelman and Stephen I. Sulzbacher
  University of Washington

- BEHAVIOR ANALYSIS & EDUCATION -- 1972 . . . . **11**

  Wesley C. Becker
  University of Oregon

- PANEL DISCUSSION: TRAINING BEHAVIOR MODIFIERS. **26**

  Jack Michael, *moderator*

  Jon Bailey, David Born, Willard Day,
  Robert P. Hawkins, Howard Sloane
  and W. Scott Wood, *discussants*

PART 2: *CURRENT TOPICS IN BEHAVIOR ANALYSIS* **35**

- BEHAVIOR ANALYSIS OF DRUG EFFECTS IN THE CLASSROOM. **37**

  Stephen I. Sulzbacher
  University of Washington

- THE EFFECT OF GENERAL & DESCRIPTIVE REINFORCEMENT ON "CREATIVITY" IN EASEL PAINTING **53**

  Elizabeth M. Goetz and Mary M. Salmonson
  University of Kansas

- FREEDOM & RESPONSIBILITY IN AN ELEMENTARY SCHOOL. **62**

  Charles L. Salzberg
  *University of Kansas
  and Colorado Springs Community School*

- ACHIEVEMENT TESTS VS. DIRECT & DAILY MEASUREMENT. **78**

  Marie D. Eaton and Tom C. Lovitt
  University of Washington

page

● ACQUISITION OF PHONETIC SOUNDS BY PRESCHOOL CHILDREN
I. EFFECTS OF RESPONSE AND REINFORCEMENT FREQUENCY
II. EFFECTS OF TACTILE DIFFERENCES IN DISCRIMINATIVE STIMULI. 88

Virginia I. Massad and Barbara C. Etzel
University of Kansas

● OPEN CLASSROOMS: SUPPORTERS OF APPLIED BEHAVIOR ANALYSIS. 112

Mary McNeil, Susan Hasazi,
Adler Muller and Martha Knight
University of Vermont
and Hinesburg Elementary School

● INTEGRATION OF NORMAL & HANDICAPPED CHILDREN IN A BEHAVIOR MODIFICATION PRESCHOOL: A CASE STUDY. 127

K. Eileen Allen, Paulette M. Benning
and W. Thomas Drummond
University of Washington

PART 3: *PROGRAMS & TECHNIQUES IN BEHAVIOR ANALYSIS: A COMPREHENSIVE BEHAVIOR ANALYSIS PROGRAM FOR HANDICAPPED CHILDREN* 143

● PRINCIPLES AND OBJECTIVES IN AN ACADEMIC PROGRAM FOR YOUNG HANDICAPPED CHILDREN 145

Jeffrey A. Grimm and Sidney W. Bijou
University of Illinois

● THE INDIVIDUALIZED READING PROGRAM 155

Linda N. Berner and Jeffrey A. Grimm
University of Illinois

● APPLICATION OF BEHAVIOR PRINCIPLES TO THE TEACHING OF WRITING, SPELLING & COMPOSITION 171

Ely Rayek and Elizabeth Nesselroad
University of Illinois

● THE RECIPROCAL MODIFICATION OF ARITHMETIC BEHAVIOR AND PROGRAM DEVELOPMENT. 185

Joseph A. Parsons
University of Illinois

page

- A FINAL NOTE **200**

  J. A. Grimm
  University of Illinois

PART 4: *PROGRAMS & TECHNIQUES IN BEHAVIOR ANALYSIS: USING CLASSROOM STUDENTS AS BEHAVIORAL ENGINEERS.* **203**

- THE EFFECTS OF A PEER CORRECTION PROCEDURE ON THE ARITHMETIC ACCURACY FOR TWO ELEMENTARY SCHOOL CHILDREN **205**

  Miriam F. Conlon, Carolyn Hall
  and Edward M. Hanley
  University of Vermont

- A BEHAVIORAL APPROACH TO REMEDIAL READING USING STUDENTS AS BEHAVIORAL ENGINEERS **211**

  Jerry Willis, Jeane Crowder and Betty Morris
  Jefferson County Department of Health
  Birmingham, Alabama

- EFFECTS OF PEER TUTORING ON THE SPELLING PERFORMANCE OF ELEMENTARY CLASSROOM STUDENTS **222**

  V. William Harris, James A. Sherman,
  David G. Henderson and Melissa S. Harris
  University of Kansas

- EFFECTS OF HAVING ONE REMEDIAL STUDENT TUTOR ANOTHER REMEDIAL STUDENT **232**

  Michael Davis
  University of Utah

PART 5: *PROGRAMS & TECHNIQUES IN BEHAVIOR ANALYSIS: TEACHER TRAINING* **251**

- SETTING CRITERION RATES OF TEACHER PRAISE: THE EFFECTS OF VIDEO TAPE FEEDBACK IN A BEHAVIOR ANALYSIS FOLLOW THROUGH CLASSROOM **253**

  Richard A. Saudargas
  Florida State University

- SCHEDULES OF FOLLOW-UP AND THEIR EFFECT UPON THE MAINTENANCE OF A PRESCRIPTIVE TEACHING PROGRAM **262**

  Beverley A. Holden and Beth Sulzer-Azaroff
  Southern Illinois University

page

● A COMPARISON OF THREE DIFFERENT TYPES OF FEEDBACK ON TEACHERS' PERFORMANCE **278**

Sarah Rule
University of Kansas
and Colorado Springs Community School

● 3 STUDENT STUDIES **290**

Patricia M. Pierson
University of Vermont

PART 6: *PROGRAMS & TECHNIQUES IN BEHAVIOR ANALYSIS: POTPOURRI* **301**

● THE EXPERIMENTAL ANALYSIS OF A HOMEWARD EDUCATIONAL PROGRAM WITH A PRE-DELINQUENT JUVENILE. **303**

Martin J. Pollack, Beth Sulzer-Azaroff
and Reed Williams
Southern Illinois University

● THE AUTO-PEDESTRIAN SAFETY PROJECT. **310**

Donald A. Jackson, William J. Mayville
and Joseph B. Cowart, Jr.
University of Utah

● A PROGRAM OF DELAYED CONSEQUENCES FOR THE MANAGEMENT OF CLASS ATTENDANCE AND DISRUPTIVE CLASSROOM BEHAVIOR OF 124 SPECIAL EDUCATION CHILDREN **318**

Hewitt B. Clark
Western Carolina Center

● 3 SIMPLE PUNISHMENT TECHNIQUES FOR THE CONTROL OF CLASSROOM DISRUPTIONS. **331**

Thomas Sajwaj, Penelope Culver,
Carolyn Hall and Les Lehr
University of Mississippi Medical Center

● USING REINFORCEMENT CONTINGENCIES & TEACHING AIDS TO ALTER SUBTRACTION PERFORMANCE OF CHILDREN WITH LEARNING DISABILITIES **342**

Deborah D. Smith, Tom C. Lovitt
and John D. Kidder
University of Washington

page

- SUPERIMPOSING CONTRACTS UPON A TOKEN ECONOMY. **361**

  Joseph R. Jenkins and Sheila Gorrafa
  New Mexico State University

PART 7: BEHAVIOR ANALYSIS IN HIGHER EDUCATION **369**

- COLLEGE STUDENT STUDY BEHAVIOR IN A PERSONALIZED INSTRUCTION COURSE AND IN A LECTURE COURSE. **371**

  David G. Born, Michael Davis,
  Paul Whelan and Donald Jackson
  University of Utah

- THE MEASUREMENT & ANALYSIS OF COLLEGE STUDENT STUDY BEHAVIOR. **377**

  James M. Johnston, Michael D. Roberts
  and George W. O'Neill
  Georgia State University

- A MULTIPLE BASELINE ACHIEVEMENT TEST. **393**

  L. Keith Miller and F. Hal Weaver
  University of Kansas

- THE EFFECT OF TWO TEST-RETEST PROCEDURES ON THE CLASSROOM PERFORMANCE OF UNDERGRADUATE COLLEGE STUDENTS. **400**

  Darrel E. Bostow and Gerald J. Blumenfeld
  University of South Florida

part 1:

BEHAVIOR ANALYSIS:

past, present, and future...

DICKY AT 13 YEARS OF AGE: A LONG-TERM SUCCESS FOLLOWING EARLY APPLICATION OF OPERANT CONDITIONING PROCEDURES[1]

deborah nedelman
stephen i. sulzbacher

When we attempt to evaluate the effectiveness of operant conditioning techniques in altering dysfunctional behavior, we will need to look for evidence of long-term success. The following case study provides such evidence.

Dicky was the subject of one of the earliest published reports of the use of operant techniques with children by Wolf, Risley and Mees (1964). He was 3 1/2 years old and in a state mental institution at the time of that study; he is thirteen years old today and is enrolled in a regular sixth grade class in public school.

Most of you will recall that original paper and the subsequent ones by Wolf, Risley, Johnston, Harris and Allen (1967) and by Baer (1966) which reported on the follow-up work done with Dicky in a preschool setting.

Figure 1 illustrates in some detail the extent of Dicky's treatment history. Each upward excursion of the line indicates an entry in his medical record. At the age of nine months bilateral cateracts were discovered, and he had a series of examinations and surgery before the age of two.

It was after this that he began to manifest the self-destructive behavior that was to bring him to the attention of behavior modifiers: banging his head, scratching his ears and neck until they bled, slapping his face, etc. At age two years two months, Dicky had his first psychiatric visit; he was diagnosed as "psychotic and possibly retarded." On the second visit a few months later, the psychiatrist noted that the mother exhibited depression and manifest anxiety, and referred her for psychiatric help. During the next few years there were several consultations with various doctors and clinics. Dicky was given several different labels, but everyone concurred that the child had severe behav-

---

[1] This research was supported by Maternal and Child Health Service Project #913 at the Child Development and Mental Retardation Center, University of Washington (Departments of Pediatrics and Psychology). Address any communications to: S. I. Sulzbacher, Pediatrics Department (WJ-10), University of Washington, Seattle 98195.

ioral problems and more than one reference is made to the mother's "depressive reaction."

When Dicky was in his fourth year, his behavior problems had progressed to such an extent that it became necessary to place him in the state mental hospital for a three month period. It was at this point that the group from the Developmental Psychology Laboratory (DPL) at the University of Washington was called in to work with Dicky. The group's efforts to control his temper tantrums, glasses-wearing and sleeping behavior were dramatically successful and are described in the previously mentioned references.

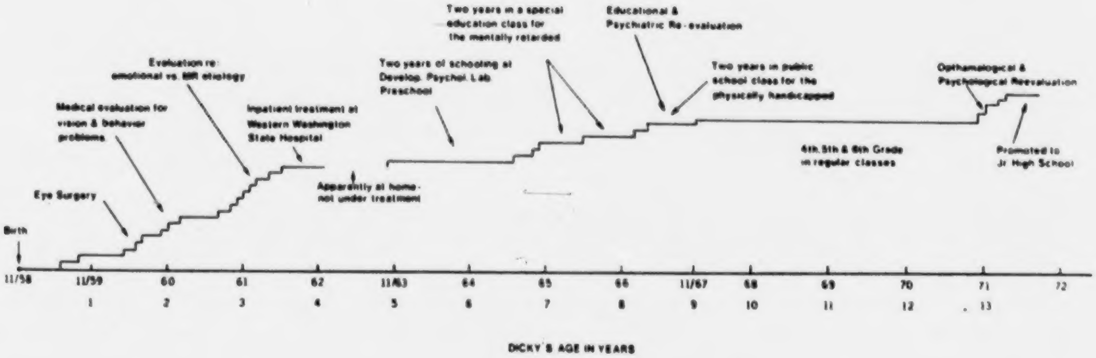


Fig. 1. Dicky's treatment history.

Upon his return home from the state hospital, Dicky was enrolled in the preschool at the Developmental Psychology Laboratory. Here much effort was made to develop Dicky's socialization skills. Since his first three years had been spent primarily in hospitals, Dicky had not had the normal opportunities to explore his environment nor to interact with peers outside of the institutional setting. Although systematic attempts to shape social behavior were not considered entirely successful by the DPL group, with neither contingent teacher attention nor contingent edible consequences significantly increasing Dicky's interactions, his rate of social interaction with his nursery school playmates did increase from near zero throughout his first year to a median of 15% of his day in his second year at the preschool.

A program of tutorial sessions was also begun at DPL by Baer and McCleave to teach Dicky to read; his behavior had to be progressively shaped to provide him with the visual discrimination skills necessary for reading. Dicky's verbal behavior was originally limited to occasional echoic responses, jargon and songs. To bring this under environmental control, Dicky was required to emit on demand the desired behavior to receive bites of his meals. He was taught to match and label letters and then to mimic names of a series of five pictures. Gradually the teacher would omit saying the word first and Dicky soon came to say the word in the presence of the picture alone. He then progressed to picture books, common household objects, and finally to events which had occurred at a previous time.

Dicky remembers this period quite vividly as was seen in a video tape shown at the conference.[2]

When he finished his two years at preschool, Dicky entered the public schools in a special education class for the mentally retarded. He adjusted well to this setting and in 1967 entered a special class for the physically handicapped. At this time Dicky was re-evaluated at the Child Development and Mental Retardation Center, University of Washington, and significant progress was found in all areas of behavior, though extreme overprotectiveness by the mother was still noted.

Dicky's ability to handle the academic world had evolved so significantly that by 1969 he was attending regular classes, and is presently enrolled in a regular 6th grade class.

Figure 2 is a graph of Dicky's IQ testing history. You will note that at the age of three he was untestable (the Vineland Score came from an interview with his mother). His current functioning IQ is 81, though his verbal score is 106, at the upper end of the average range. From the age of 7 1/2 to 13 he gained fourteen points on his verbal score. The drop of twelve points in his performance score during this same period reflects his still quite severe visual handicap.

---

[2] On February 2, 1972 at the Child Development and Mental Retardation Center at the University of Washington, a video tape was made of a conversation between Dicky and Mrs. Eileen Allen and the present authors. The tape was shown as part of this presentation.

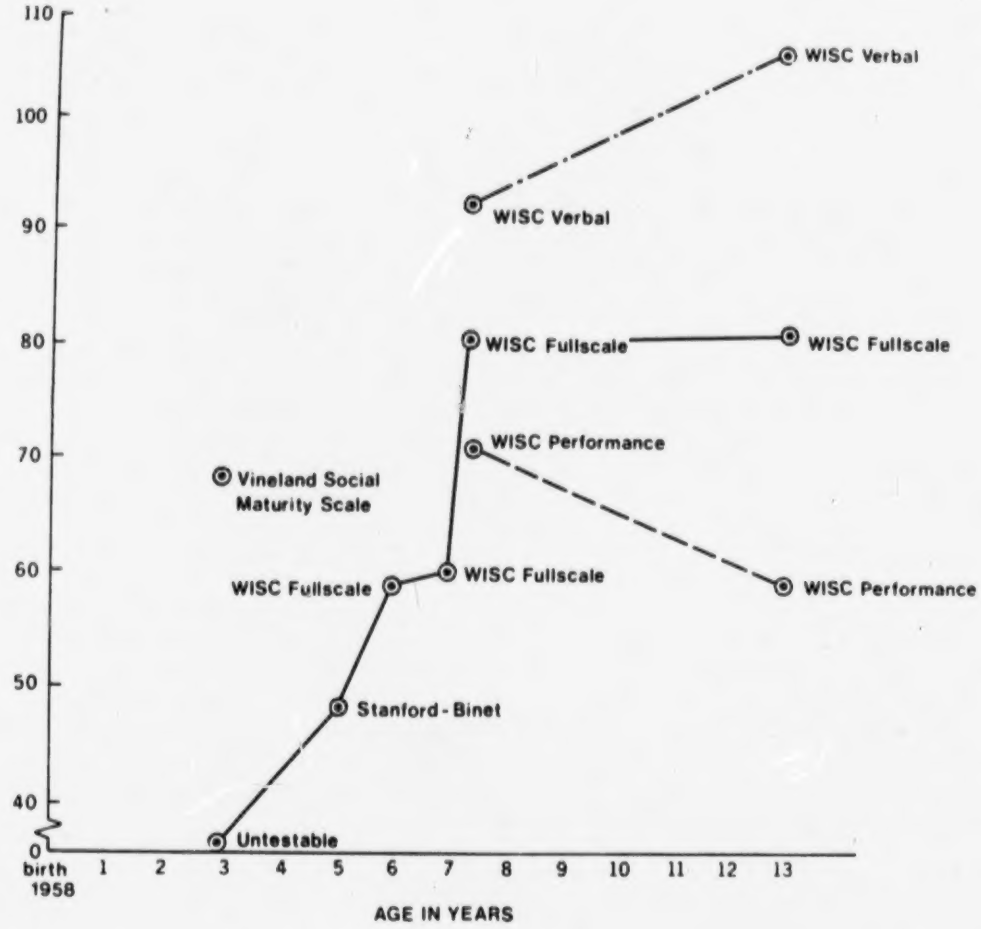


Fig. 2. Summary of Dicky's intelligence testing performance.

As part of Dicky's psychological evaluation, projective tests were administered in 1967 and again in February, 1972. While some may question the validity of these tests, per se, it is worth noting that the improvement in this child's behavior is correlated with a "healthier" profile on these tests. Some of the specifics noted by the psychologist were: (with reference to the Rorschach) "Whereas his previous record was *flexible to the point of no boundaries to his concepts* now all of his responses appear to be reasonably well related to the inkblots;" (with reference to the TAT) "He can now describe what the characters are doing rather than merely listing the items in the picture. He carefully describes facial expressions such as *straight-faced, love-expression, happy,* as if he had been coached to do this in social situations." The conclusion reached by the tester was: "On the whole, projective tests indicate remarkable improvement and suggest that Dicky is at present content with himself and his situation in life. Although he is still decidedly dependent and immature for thirteen years, he is not overly anxious, frustrated or threatened."

Dicky's steady progress has brought him to the point where he is regarded by his current teacher as a kind and considerate student, well-liked by his classmates; a student who works quietly, pays attention in class and takes pride in his work. His reading level is appropriate for his grade. Most of his reading material is provided in large print, though he can read some regular-sized print, as he mentions in the film. He is receiving satisfactory grades in all subjects except for math and handwriting. He has recently begun working with a tutor in math, where he is having some difficulty with long division.

On a recent trip to his school, Dicky was observed to be generally socially appropriate and particularly sensitive to the feelings of other children who were having difficulty. At a student government meeting where he represents his homeroom, Dicky made a plea for greater understanding of the behavior of an underprivileged child who had been the object of petty persecution by other students.

Dicky still displays some "autisms" or "blindisms:" rocking in his seat and occasionally hand flapping, though his classmates seem to attribute this behavior to his visual handicap. Dicky was observed at recess in 1967 and in 1972 and on both occasions did not engage in any play activity with other children, but his teacher was surprised by this and said that it was not common for him any longer. One explanation for his isolate play at recess is that his poor vision would interfere with any ball games, which were the predominant activity.

Dicky will be entering a regular junior high school class next fall.

Dicky's ability to participate in a prolonged conversation and his appropriate use of a relatively sophisticated vocabulary were evident when we talked with him. The conversation covered many diverse areas, from his memories of nursery school where Mrs. Allen was his teacher, to his current bowling score. Dicky talked about himself as being "handicapped" and by this, as he explained, he meant visually impaired. Here are some quotes from the video tape of that conversation:

✱*(Discussing an earthquake that had occurred when Dicky was in the DPL preschool.)*

E^1: Tell me about it, Dicky, I wasn't there.

D : It was a pretty bad earthquake. I was in the building..... All of a sudden without the slightest warning it happened. Everything got knocked over..... 'cause the vibration was so great.

✱*(Discussing his current friends.)*

D : Mark and Larry and Greg are my best friends. Greg is kind of handicapped, too.

E^2: In what way?

D : He has a problem talking just like I have a problem seeing. We don't have too many handicapped kids in our class.

E^2: Any others?

D : ....Well, probably about a fourth of the class wears glasses.

✱*(Discussing his impairment.)*

D : ...I could see good when I was a baby, but by the time I was about eight or nine months old I started having trouble with my eyes--cataracts. I was almost blind then.

E^2: I remember, you had some operations on your eyes.

D : Five. I was almost blind. I couldn't hardly see anything.

E^2: When you were about 3 or 3 1/2 you started learning to wear your glasses.

D : Yes. But my eyes got a little better...

E^2: ...Can you read regular print now, or do you have to have special print?

D : I can read regular print to a certain extent. I can't read small print, newspaper print.

DISCUSSION

It can be seen that the prediction made by Wolf and his associates that Dicky would become increasingly verbal and "a new source of joy to the members of his family" and, "perhaps, through the continued efforts of his parents and teachers he may develop into a productive citizen," is coming true. Some remnants of his "autistic" behaviors are still evident to the trained observer, but neither his teachers nor his peers seem to notice.

There is no evidence that any new deviant behaviors ever were "substituted" for those which were the object of his early treatment. However, Dicky has been extremely fortunate in having had exemplary teachers who have continued to use classroom procedures compatible with those which were initially so successful. The case of Dicky supports the contention by Lovaas (1968) and Sulzbacher and Costello (1970) that the lives of seriously handicapped children can be materially enriched by early intervention using the behavior analysis model and follow-up over a period of years to insure generalization and continued acquisition of new skills.

REFERENCES

Baer, D. M. An age-irrelevant concept of development. Paper presented at American Psychological Association, New York, 1966.

Lovaas, O. I. Some studies on the treatment of childhood schizophrenia. *Research in Psychotherapy*, 1968, 3, 103-121.

Sulzbacher, S. I. and Costello, J. M. A behavioral strategy for language training of a child with autistic behaviors. *Journal of Speech and Hearing Disorders*, 1970, 35, 256-276.

Wolf, M. M., Risley, T. R., Johnston, M., Harris, F. and Allen, E. Application of operant conditioning procedures to the behavior problems of an autistic child: a follow-up and extension. *Behavior Research and Therapy*, 1967, 5, 103-111.

Wolf, M. M., Risley, T. R. and Mees, H. L. Application of operant conditioning procedures to the behavior problems of an autistic child. *Behavior Research and Therapy*, 1964, 1, 305-312.

BEHAVIOR ANALYSIS & EDUCATION --1972

wesley c. becker

As I look at the progress of behavioral approaches in psychology and education over the past decade, I am really amazed at the impact a few people like Staats, Michael, Wolf, Bijou and Baer have been able to make on research methods in psychology and applications to social problems. Most of us at this conference are testimonials to their influence.

The decade has produced many procedures for eliminating or changing behavior problems in a variety of settings. The decade has produced programs for training parents, teachers, and college students in the basics of behavioral analysis and its application in controlling problem behavior. A small beginning has been made in the building of instructional systems using behavior knowledge. We are practically nowhere in working out applications to the administration of educational systems.

As a person out in the field of public education, I see that we are at a critical point where whether we stand still for the next decade or have the chance to make an impact on education at all levels is going to depend on three things:

1▶ gaining a favorable public press and political support for behavior analysis;

2▶ producing educational systems that teach efficiently and positively, and data to show this; and

3▶ getting a foothold on teacher training at the university level, and producing procedures for training performance competencies which affect child learning.

There is currently on Elliot Richardson's desk, the Hively report on applications of behavior analysis to education. If acted on, groups of us could very readily organize to: (1) implement field use of current knowledge on applications of consequences in schools, and (2) produce the next generation of research demonstrations required on teacher training and on instructional systems.

A better public press and active political support is required, however, to bust this report loose in the face of criticism of behaviorists by uninformed politicians and journalists. Similarly, the directions taken by a National Institute of Education, if it ever gets going, are right now critically sensitive to any public impacts we can make in the near future.

I am amazed at the disservice some behaviorists do to each other in explaining themselves to novice students and to the lay public. Questions of control, of responsibility, and of freedom have been poorly handled in most cases. Skinner's latest book, *Beyond Freedom and Dignity*, has been misinterpreted because it fails to present the issues in a way that deals adequately with current belief and expectations. Besides the Agnew attack, Skinner has been attacked on the floor of the House by Cornelius Gallagher for "advancing ideas which threaten the future of our system of government." Well, we know this is not true--but the attacks cannot be ignored. I have read many manuscripts in the past year where behaviorists are trying to explain to teachers or college students why man is not free and not responsible, and that these things are just illusions. It doesn't do much good to assert this without explaining the "illusion." I think that there are some better ways of dealing with these issues.

Allow me to get didactic for a few minutes, and suggest one analysis which deals with the problems of freedom, control, and responsibility in a way which I think we can sell to most reasonable human beings. I know Dick Mallot could do a more creative job of getting these ideas across, and I hope he will give it a try.

Do Behaviorists Deny That Man is Free?

There are two answers to this question:

The First Answer. If by *free* we mean that man has a free will and can *autonomously* decide for himself what he will do independently of influences upon him from his environment, then behaviorists must assert that man is not free. It is quite clear from experimental work that behavior is controlled by environmental events. This statement, however, is equivalent to saying that people learn what they are taught by their environment. Notice how readily some could accept this latter statement, but still want to deny that behavior is controlled by our history. In any case, it is this concept of freedom which Skinner is asserting to be illusory. That we learn what we are taught should be obvious. Few of us speak Russian or French or German, because we were not taught those languages by our parents.

The Second Answer. If by *free* we mean that man frequently faces choice situations and his political system allows him to make those choices according to his own history of experience, without *coercion*, then man can be free. In choice situations, men choose the more rewarding and less punishing alternative as determined by their prior experiences. Choice behavior is determined like any other behavior. The *issue* of freedom of choice arises when *other people* have power to force us to make choices we would not choose because of our prior histories. When a parent forces a child to go to bed, but the child wants to watch TV, the child is experiencing a loss of freedom. When the state passes a law requiring manufacturers to stop polluting waterways under threat of penalty, the manufacturers are losing a freedom of action they once had. When a man signs a marriage contract with a woman, he is entering into a legal contract that usually restricts the choices he can make of sexual partners.

I realize most of you understand these things, but indulge me a little more. Through centuries of philosophic thought, confusions have existed in understanding freedom of choice and lawfulness in behavior. The confusions can be largely attributed to confusions in the meanings of words. Consider the following concepts:

"Natural Laws"[1]	*Man-Made Laws*
Describe what is	Prescribe what should be
Determinism vs. Chance	Coercion vs. Freedom of Choice

Much misunderstanding can be avoided if we first clearly distinguish between natural laws and man-made laws. Natural laws simply describe the way we understand things to be in nature--what leads to what. For example, preceding stimulus events control reflexes; or, water boils when heated to 212° at sea level. As some wag put it, "nature doesn't violate her own laws," and being part of nature, neither do we. We can't. That's just the way it is. Theoretically, natural laws are true whether we like them or not, whether we know them or not. Man-made laws on the other hand, prescribe what you should do and the consequences for not doing something. Men make them and men can change them. They are not true, but only potentially useful. All explicit or implicit contracts and agreements between people have a similar prescribing function. The obligations of people to each other are what defines a social system, a society.

With the notion of natural law versus man-made law clearly in mind, let's consider the other concepts listed above. In science, determinism is the working hypothesis that there is lawfulness (consistency) in nature to be discovered. The opposite of determinism is chance. If something is not orderly and consistent, then it is random. Throughout philosophic history[2], it has been asserted that freedom is the opposite of determinism, and, therefore, if one is true the other cannot be. But, this is not the case. In the natural world the opposite of determinism is chance, and in the man-made societies the opposite of freedom of choice is being coerced or forced to do what you would not do if the coercion were not present. Thus, if we clearly examine the meanings of the words involved in this issue it is not hard to come to a reasonable resolution. Behaviorists can readily agree that some forms of social systems involve more freedom of choice for the people than others. Most behaviorists would go further and say that with our current knowledge, it is possible to design even better social systems where there is less need for coercion (punishment), less misery, less threat of future destruction of mankind; and more individual choices, more happiness, more accomplishments, and more appreciation of living.

---

[1] Of course *"natural laws"* are man-made statements, but the processes referred to are not man-made.

Do Behaviorists Deny That Man is Responsible?

If man's behavior is determined, how can man be considered responsible for his actions? As with freedom, the question is based on a misunderstanding. Responsibility is *assigned* to people by man-made laws and contracts. For example, laws of state make parents responsible for their children, not laws of nature. The assignment of responsibility by man-made laws and contracts, does not necessarily make a man responsible. It simply specifies the penalties for failure to live up to a responsibility or the rewards for doing so. Many are punished for failing to live up to their responsibilities. Man is law abiding and responsible if, as Chuck Salzberg (1972) pointed out this morning, he has been trained to be so by others in his society. The current knowledge from behavior theory provides parents and others with the knowledge needed to train children to be responsible citizens. Behaviorists need not deny the validity of contractual systems and the responsibilities imposed by them; in fact, they are in a position to show the world how to build better contractual systems through empirical analysis of effects.

The dilemma we are in has in part been generated by our zeal to extol the virtues of functional analysis, and to put down hypothetical-deductive research. But, the effect has been to fail to teach our students that there is a time for thinking about what you are doing, a time for logical analysis in considering what we know and where we can go with it. We have failed to teach our students to distinguish between evaluation which is functional and design which is procedural.[3] Note that nearly all of the papers in this conference deal with the evaluation of comparative designs. The functional-procedural distinction underlies the difference between basic and applied science.

---

[2] See Schlick, Moritz. *Problems of ethics*. New York: Prentice-Hall, Inc., 1939.

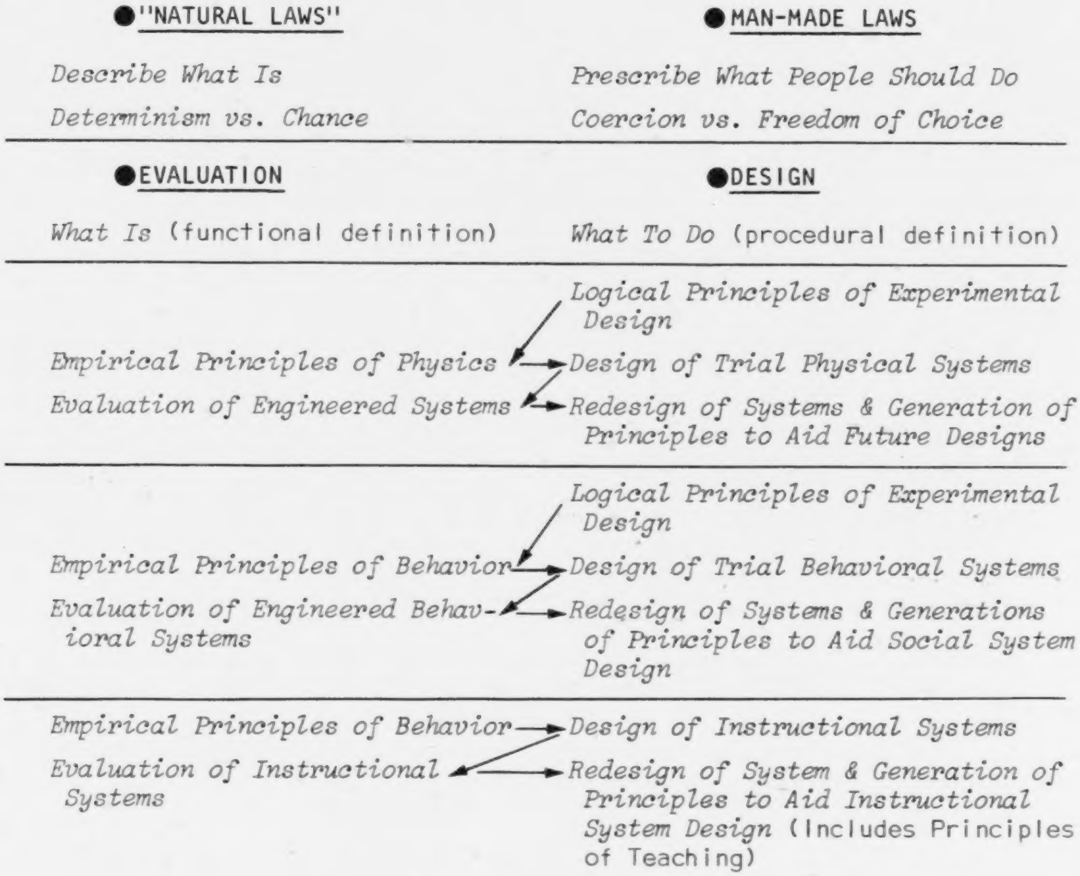

●"NATURAL LAWS"
Describe What Is
Determinism vs. Chance
●MAN-MADE LAWS
Prescribe What People Should Do
Coercion vs. Freedom of Choice
●EVALUATION
What Is (functional definition)
●DESIGN
What To Do (procedural definition)
Logical Principles of Experimental Design
Empirical Principles of Physics
Design of Trial Physical Systems
Evaluation of Engineered Systems
Redesign of Systems & Generation of Principles to Aid Future Designs
Logical Principles of Experimental Design
Empirical Principles of Behavior
Design of Trial Behavioral Systems
Evaluation of Engineered Behavioral Systems
Redesign of Systems & Generations of Principles to Aid Social System Design
Empirical Principles of Behavior
Design of Instructional Systems
Evaluation of Instructional Systems
Redesign of System & Generation of Principles to Aid Instructional System Design (Includes Principles of Teaching)

Fig. 1.

---

[3] For the novice to behavior theory, functional analysis involves the controlled demonstration that certain environmental events affect behavior in certain ways. Functional definitions specify an environmental event, a procedure for using this event, and an effect on behavior. (It *is* a reinforcer if it follows behavior and strengthens its future rate of occurrence.) The design of systems is a logical process in which procedural definitions are used. For behavioral systems procedural definitions would specify environmental events and how to use them in relation to behavior, but not specify the outcome or behavioral effect. ("We are going to use these tokens and these back-ups as reinforcers in our system.")

Refer to Figure 1 in considering the following discussion. Functional evaluation procedures are used to find out what is. Logic is used to design procedures for operating on the environment. In the laboratory we start with the logical requirements of an experiment. How can we find out what is going on? One rule is to change only one thing at a time. Another is to repeat the procedures to be sure the effect is real. Experimental designs are used to find out about functional relations. In the physical world, these relationships are called the principles of physics. These principles are then applied in the design of man-made physical systems. This is called engineering. Systems once designed are evaluated for their functional characteristics just as "natural" systems were. Evaluation then is followed by redesign and the generation of principles of engineering. Recycling can occur here. Note that the principles of engineering are not true, but only useful. Someone could design systems that work better tomorrow and produce new design principles.

A similar interaction of logical and empirical processes follows in the behavioral world. Logic guides experimental design. Empirical generalizations are found and applied to trial behavioral systems. These are evaluated and redesigned. Principles of social engineering may also be an outcome. The kinds of questions answered might be, "Who should control the reinforcers or punishers in a community like Achievement Place?" "What should be used as reinforcers for these drug addicts?" "How should the contracting arrangements between teachers and high school students be arrived at?" The basic question is always which procedures are more effective under given conditions, given a criterion of utility?

Let me return now to where we are going. My major point in focusing on these conceptual distinctions is that we all need to be smarter in coping with unwarranted attacks. We need to be able to show those who attack on the wrong basis just where they are misunderstanding what we are doing. We need to be better teachers, and we need to be political about it, or we may be in for repressive attacks.

Now this analysis I have given you, just happens to also lay out for you a framework within which to consider where we should be going. What are we shooting for? As I said earlier, I see a need now to produce demonstrations showing how we can use our growing sophistication to produce efficient educational systems, including educational environments, and to improve teacher training. We need to be designing and evaluating the behavioral systems which will someday provide the principles for the field of social engineering.

Those of us working in Follow Through (Bushell, Engelmann, and myself) have been trying to develop such systems and demonstrations. Let me just give you a very brief summary of the system we are using in twenty school districts with 10,000 disadvantaged children.

The Engelmann-Becker Follow Through Model

The goal of our model is to teach poor kids more, not less. There are six basic components to the system we devised:

●1. *We increased the manpower in the classroom by adding two aides.*

●2. *We structured the daily program so that the teaching personnel had a clear plan of action.*

●3. *We provided daily programmed lessons to insure that the teaching personnel knew what was to be taught in what sequence.*

●4. *We instituted a teaching method that utilizes basic behavioral principles to insure efficiency in teaching.*

●5. *We provided for training in the use of the programs on a continuing basis.*

●6. *We monitored the progress of the children and the skills of the teachers and aides to be sure the system was functioning.*

The keys to this model which are different from other current behavioral instructional systems are: (1) *the ways in which Engelmann has programmed the instructional materials to teach a general case. This programming focuses analysis on concepts and operations needed at higher levels of the program, rather than task analysis, in making design decisions;* (2) *the packaging of the instructional materials into daily lessons to be given by an adult working with small groups of children in direct instruction.*[4]

Figure 2 provides an example of material from a teacher's manual (now published under the trade name *DISTAR*, by SRA). Each lesson day is broken down into a number of tasks. The visual materials the teacher needs for that task are provided, or she is instructed what to draw on a chalkboard. She is given further instructions on what she should do (in black) and what she should say (in color). The required child responses are given in italics. Aides who are parents of poor children can learn to teach with these materials with two weeks preservice training, two hours a week inservice, and periodic supervision in the classroom.

Efficient teaching aims at getting a high rate of correct child responses within the teaching time available. To accomplish this, the teacher needs to know the formats in the program well. She needs to know how to use attention signals most appropriate for each task and must learn to use a

---

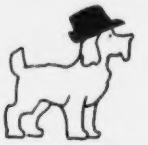
[4] See Becker, Wesley, C., Engelmann, Siegfried and Thomas, Don, 1971.

Praise the children for correct responses. Correct mistakes immediately.

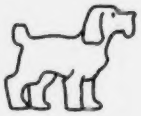
TASK 2

Group Activity

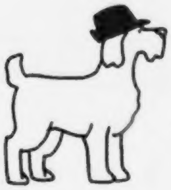
a. **Point to the dogs.** What are these? *These are dogs.* **Call on one child.** Find the dogs that are little. Everybody tell me about these dogs. **Repeat using the dogs that are wearing a hat, are big, and are not wearing a hat.**



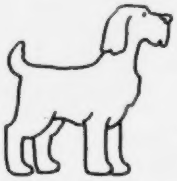
b. **Point.** Is this dog little? *Yes.* Is this dog wearing a hat? *Yes.* Is this dog little and wearing a hat? *Yes.* Say the whole thing. *This dog is little and wearing a hat.*



c. **Point.** Is this dog little? *Yes.* Is this dog wearing a hat? *No.* Is this dog little and wearing a hat? *No.* Say the whole thing. *This dog is not little and wearing a hat.* Why not? *This dog is not wearing a hat.*



d. **Point.** Is this dog little? *No.* Is this dog wearing a hat? *Yes.* Is this dog little and wearing a hat? *No.* Say the whole thing. *This dog is not little and wearing a hat.* Why not? *This dog is not little.*



e. **Point.** Is this dog little? *No.* Is this dog wearing a hat? *No.* Is this dog little and wearing a hat? *No.* Say the whole thing. *This dog is not little and wearing a hat.* Why not? *This dog is not little and not wearing a hat.*

Individual Activity

f. **Call on individual children. Point to a picture.** Is this dog little and wearing a hat? **If the dog is not little and wearing a hat, ask:** Why not?
Correction Procedure: If children answer "Is he little?" by saying, "No, he is big." say: Yes, but is he little? **If children have trouble, point to a dog.** Is he little? Is he wearing a hat? **When a child says "No." say:** Well, that's why he's not little and wearing a hat. He is not little (or) wearing a hat.

Fig. 2. An example of teacher material from the DISTAR Language I Program.

variety of signals to get the children to respond together (or individually) on cue. These latter signals we call "do it" signals, since they tell the children it is their turn to "do it." In small group instruction, "do it" signals are critical in being sure each child learns what the teacher is teaching, rather than just imitating what another child is saying. The teacher also needs to learn how to pace each task appropriately, quickly enough to hold attention, yet going slowly when required to give the children "time to think."

The teacher needs to learn to use reinforcers effectively to strengthen correct responding and to correct mistakes in a way which permits all children to learn each task (criterion teaching). Efficient correction procedures are based on an analysis of the kinds of errors which can be made. For elementary tasks, there are just three, the corrections for which are easily taught. The main point we wish to make through this discussion of training is that within a behavioral instructional system, the performance requirements for teachers (including aides) can be specified and thus training procedures can be devised. Trainers know what to do, teachers know what to do, and children learn.

Some Illustrative Outcomes

One interesting outcome is a demonstration that teaching makes a difference. There is very little data in education today which show this. Figure 3 shows classroom means for scores on WRAT reading plotted against classroom means for final day in program for *DISTAR* reading. The figure is based on data from 3,700 children. This graphic display indicates a very close relationship between the number of lessons taught to children and their test scores. The plot for reading is a consistently linear function. This function implies that efforts to ensure that more lessons are taught each child during the school year will lead to improved rates of skill development.

Figure 4 gives two-year gains for different grade groups in reading. To save time, the children for whom there was no data in 1972 were not eliminated from the previously calculated gains. Thus, there is some distortion in viewing the data as pure gain scores. However, the findings are powerful. With a little extrapolation, it can be seen that the gain in reading from pre-kindergarten to post-third grade is 4.8 grade levels. The gain from pre-first grade to post-third grade is 4.1 grade levels. Keep in mind that the expected gain for economically disadvantaged children has typically been reported at about .6 grade levels per year. Also note that the gains by children starting in kindergarten are nearly the same as those by children starting in first grade. This evidence strongly contradicts opinions that kindergarten children are not ready to learn to read, and the data are on *disadvantaged kindergarten* children.

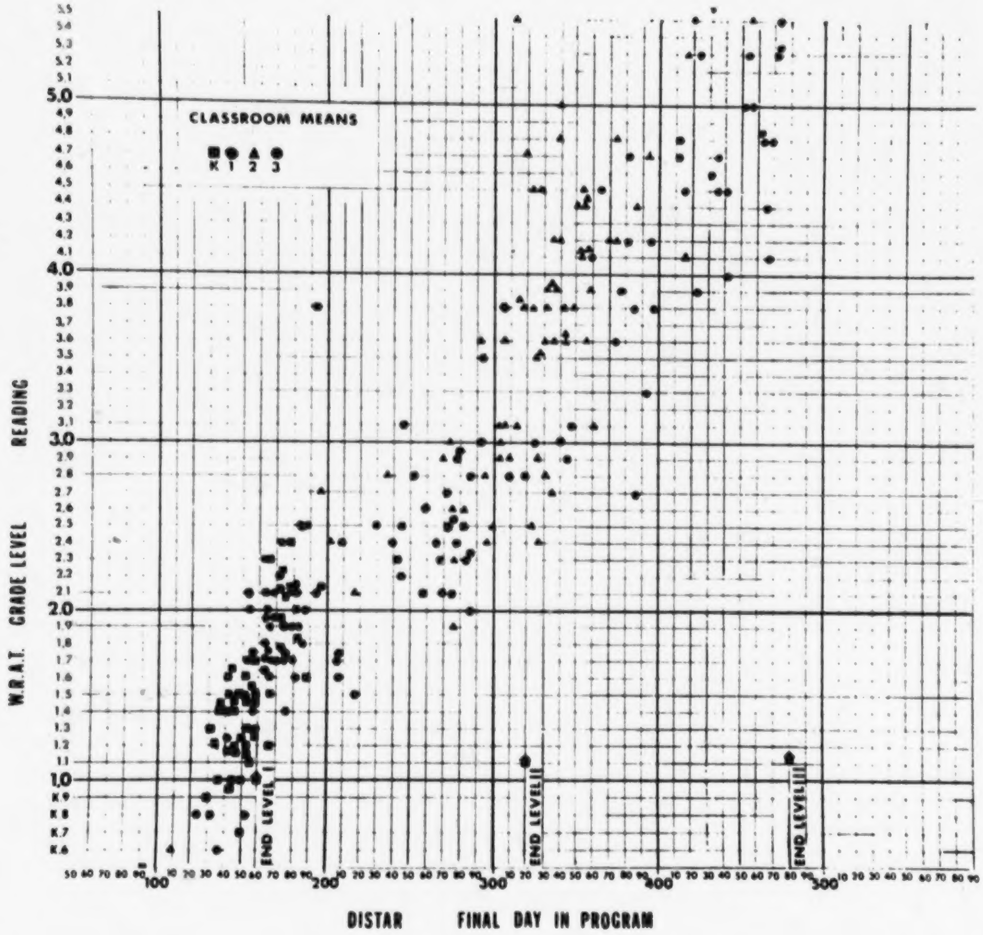


Fig. 3. Relationship of average final day in program, DISTAR Reading, to Wide Range Reading performance. Based on 3,700 children. Each dot is a classroom mean.

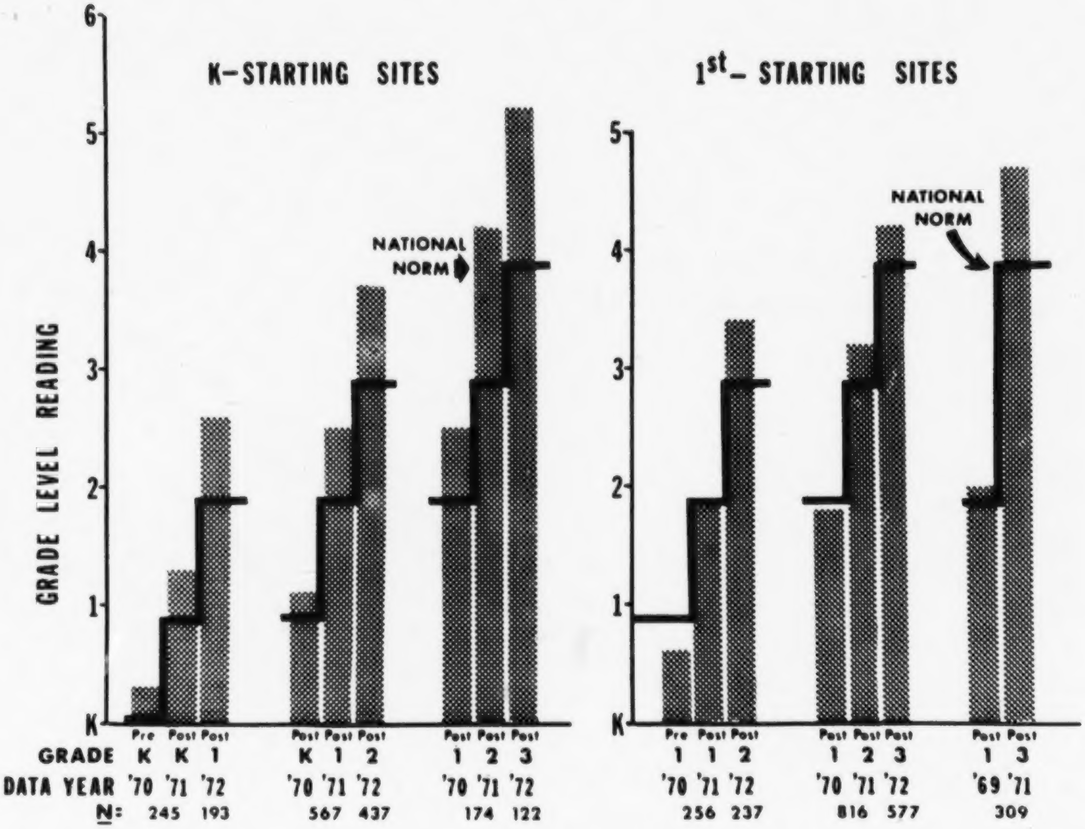


Fig 4. Wide Range Achievement Test Reading scores for poor children in the program for at least 130 days per year.

Figure 5 plots the gain scores from Figure 4 against lesson days taught in *DISTAR* reading for the periods of the gain. The figure shows a close functional relationship between gains in lessons taught and gains in WRAT reading performance. Whether the children start in first grade or kindergarten can be seen to have little effect on gain scores (parallel slopes), but children starting in first grade tend to start with more skills.

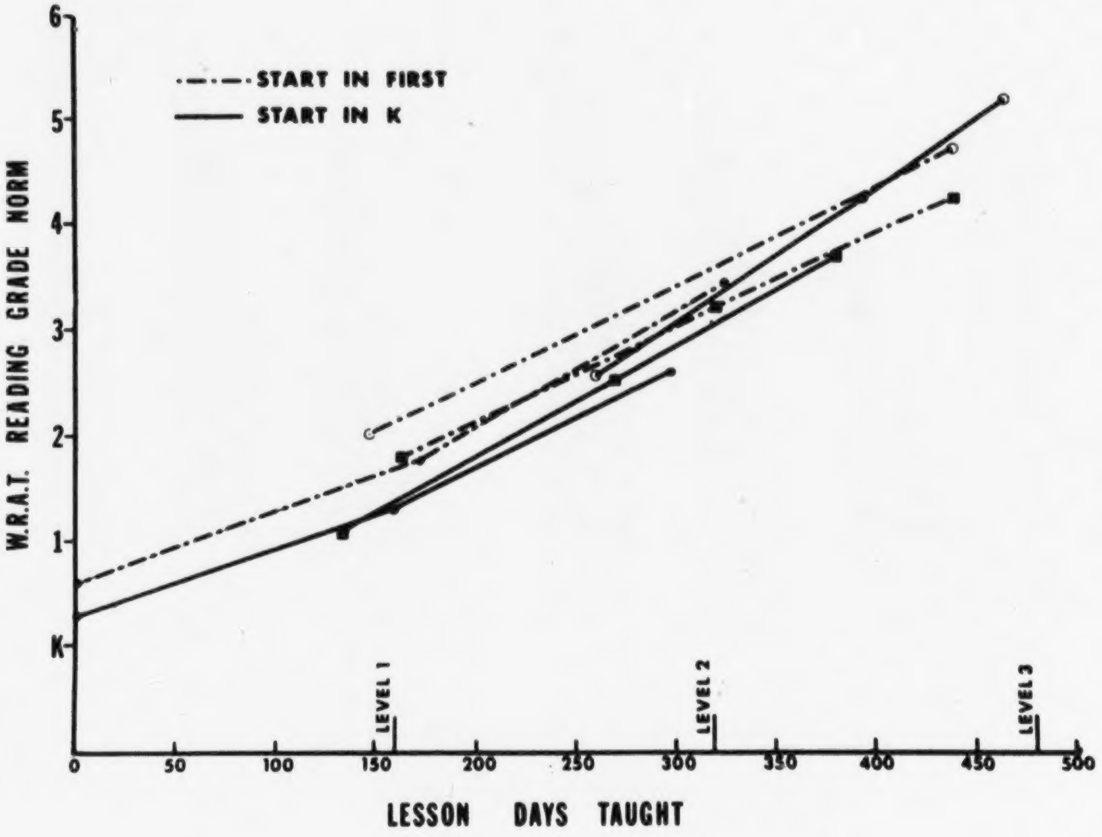


Fig. 5. Reading gains on the Wide Range as a function of lessons taught. The data are for the same six groups presented in Figure 4.

The important benefits of starting reading earlier is further supported by the analysis in Figure 6. Comparisons are made at the end of first grade and at the end of second grade between groups of children who had our kindergarten program (E-B KG.) and those who had a traditional kindergarten or no kindergarten. At the end of the first grade, those children with our kindergarten program were .7 grade levels ahead of the other two groups. At the end of second grade, these same children were one year ahead of the other two groups and more than one year above grade norms. This kind of evidence shows that just any kind of kindergarten experience will not necessarily contribute to the development of reading skills. It also shows that one way to give disadvantaged children a running start in skills they will need to be successful in school is to start teaching them earlier.

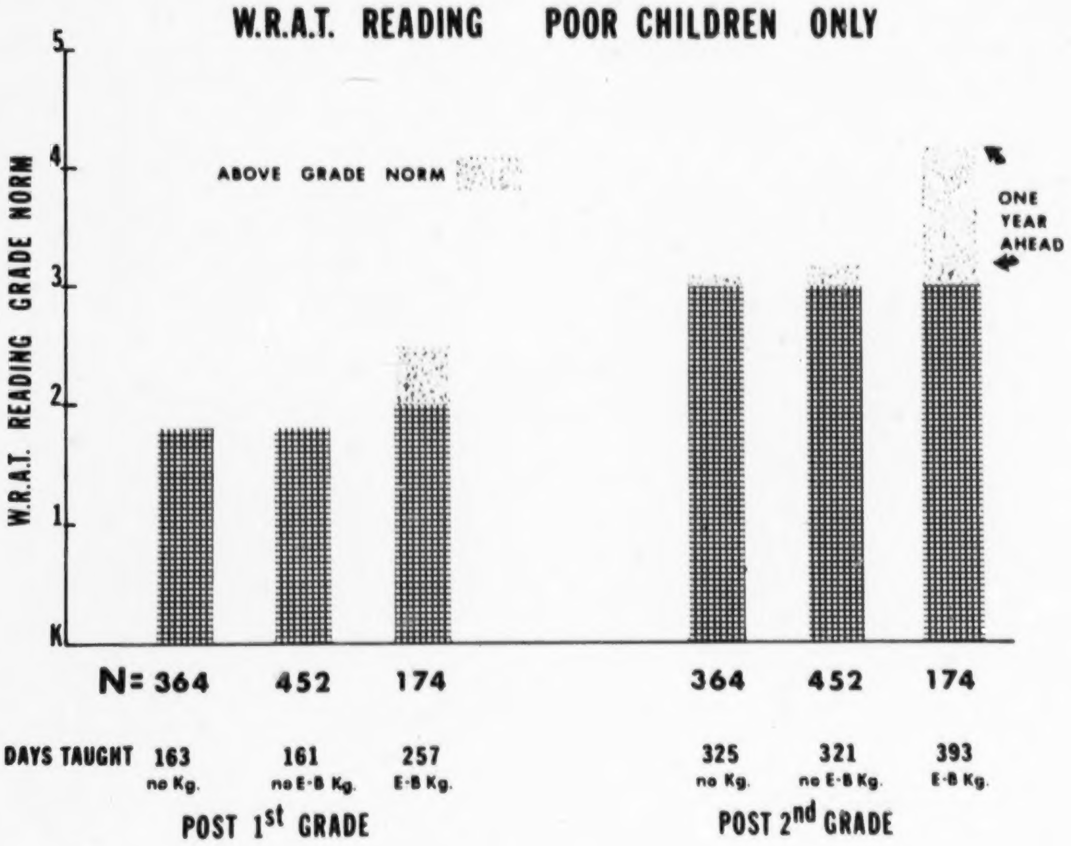


Fig. 6. Analysis of the effects of no kindergarten, kindergarten but not Engelmann-Becker kindergarten, and Engelmann-Becker kindergarten on reading levels at the end of first and second grade.

These findings are just a beginning--in showing what can be done. We have the potential for making a great impact on education and society more generally in the coming years. There is going to be a need for more of us working together to gain public acceptance and understanding, and testing out the properties of various kinds of behavioral systems on a large scale basis.

I want to close by thanking the Department of Human Development, the Follow Through Project, and Don Bushell, Jr. for hosting an exciting conference.

REFERENCES

Becker, Wesley, C., Engelmann, Siegfried and Thomas, Don. *Teaching a course in applied psychology*. Chicago: Science Research Associates, 1971.

Salzberg, Charles L. The development of freedom and responsibility in elementary school. In Semb, G., Green, D. R., Michael, J., Hawkins, R., Phillips, E. L., Sloane, H. and Sherman, J. A. (Eds.), *Behavior analysis and education, 1972*. Lawrence, Kansas: The University of Kansas Support and Development Center for Follow Through, 1972.

Schlick, Moritz. *Problems of ethics*. New York: Prentice-Hall, Inc., 1939.

PANEL DISCUSSION: TRAINING BEHAVIOR MODIFIERS

☆ **jack michael,** MODERATOR
Western Michigan University

☆ **jon bailey**
Florida State University

☆ **david born**
University of Utah

☆ **willard day**
University of Nevada

☆ **robert p. hawkins**
Western Michigan University

☆ **howard sloane**
University of Utah

☆ **w. scott wood**
Drake University

•JACK MICHAEL:

There are a number of issues we want to consider, but we haven't planned anything very organized. Scott Wood and I went over several problems regarding training and behavior modification that we thought would be useful to consider. One problem is the pressure from consumers for some type of professionalism. Now that we have an applied science, we are beginning to come under new pressures. At the same time, many of us have not been very sympathetic with the traditional clinical psychologist's concern for standardized training programs and licensing. Clinical psychologists have tried hard to decrease the likelihood that someone could pass himself off as a real clinical psychologist when all he really knew was astrology. Now the shoe is on the other foot. People frequently claim to be "behavior modifiers," but they are really not qualified. We are beginning to feel the pressure for standards, licensing, standardized training procedures and the use of labels. The problem is that there are always people present who will call themselves whatever seems useful at the moment, and who will offer services that they cannot provide. Already, a number of us have received inquiries from agencies about how to decide who to hire, which projects to fund and which services to support. Will the REAL behavior modifier please stand up?

●JON BAILEY:

In Florida, we recently had a case which indicates a real need for some kind of monitoring of behavior modification programs. A psychologist who worked for a large institution for the retarded in Miami claimed to be a "behavior modifier." As it turned out, the program he set up was simply a severe case of abnormality on the staff's part in the NAME of behavior modification. For example, a child who was caught stealing something would be forced to wear a sign, front and back, saying, "I AM A THIEF." Everyone would address him as a thief. Children caught engaging in homosexual activities were forced to re-engage in those activities in front of other staff members and patients. And, this was carried out in the name of behavior modification. Local news media reacted immediately. I think this case is an example of the issue being discussed.

Todd Risley was a member of a panel appointed to investigate the incident. We owe a great tribute to him for pointing out how this was not a behavior modification program, but rather how it involved bizarre people doing bizarre things to retarded children. His efforts resulted in some very clear statements about the use of behavior modification in retardation. For example, he emphasized the notion that you reinforce behavior that you want and that you punish behavior that you do not want, but that in the present case, the "punishment" was completely out of line. There was definitely something inadequate and inappropriate about the training those staff members received. Todd, would you want to make some comments?

●TODD RISLEY:

This is a case I would not really discuss except that the state of Florida has already made public the confidential report that the committee made. It was a rare opportunity to examine all the workings of a mental institution, to have access to all of the information and to be able to interview anybody in the institution. So, I learned a great deal. Let me characterize what happened.

The psychologist involved had credentials, including the Ph.D. They were, in most respects, good credentials for doing something other than what he was hired to do. In the areas in which he had training, he did an exemplary job. However, he had responsibility for which he had no training--dealing with juvenile delinquent kids who happened to score 69 on an IQ test and who instead of getting sent to jail got sent to the institution. They were big kids, problem kids, potentially violent and dangerous kids. In the absence of any training dealing specifically with this population, he built an upside down token economy in which the worse the behavior, the more remote the consequences. As a result of undesirable behavior, the child was sent to a ward where facilities, food and everything else was unsatisfactory. He then had to accumulate a large total number of token points to get out of this ward, and into the next ward. He could not buy anything; all he could do was save them until he had accumulated a very large number and then he could buy his way into the next program where he could finally spend the tokens he had earned. Now you can begin to see what would happen. Once a child fell out of the top system, he fell all the way out. He then got into a system where the contingencies were much more remote and delayed. In the context of this program, they had a lot of problems with the youths on the lower programs and the problems involved big kids, many of whom the staff feared.

The psychologist had no effective technology for dealing with this situation. He had some general rules, like "emphasize the natural consequences of behavior," a general statement which makes sense in an abstract way, but very little practical sense without a good deal of elaboration. Another general strategy was, "in the absence of a prescription for behavior, do something, try something--it will be reviewed later." When the staff tried something and it seemed to fit the first rule, it would be approved, or at least, not disapproved. Another rule was that you had to "follow up on any contingency statement that you make." That meant any threat or promise you made you had to back up. These rules make sense, but only in the context of a technology for dealing with the problem. Let me give you an example of how these three rules could be used. (1) Emphasize the natural consequences of the behavior: a resident was seen masturbating, so this was defined as masturbation in public, an inappropriate form of behavior in our culture. (2) Now attempt to emphasize the natural consequences of the behavior by calling other people's attention to the act in public. (I am not making this example up; this was the approximate sequence of events according to the cottage log books.) When an attendant caught a child masturbating, he drew all the other kids' attention--the "natural consequences" of public masturbation is public ridicule. The next time the resident was "caught," he stopped immediately and there was no opportunity to call everybody around. (3) However, the attendant had previously announced that when anyone was seen masturbating in public, everyone would laugh at him--and this contingency statement must be followed up. So, perhaps it seems reasonable to coerce him into re-engaging in the act and then call everyone to ridicule him. This procedure escalated in regular steps such that the resident stripped and re-engaged in public masturbation in front of other residents and staff. You can see a sort of logical progression in an isolated ward where there was no feedback or checks, no peer review, no interaction with the larger community. The basic thing is that the psychologist came into the situation with no prescriptive program. He had no technology. He had been trained to deal with principles or even with behavior problems in other populations in other places. But he had no technology for dealing with those kinds of problems. Furthermore, there was no way those who hired him could know he had not been trained to deal with these problem behaviors. This is probably a good example to deal with.

●JACK MICHAEL:

There seem to be several options at this point. One approach is the one that other professionals in psychology have adopted--they agitate for state licensing, set forth generally agreed upon standards that constitute appropriate training, and attempt to protect the title by making it unlawful to label yourself as a "behavior modifier" unless you have been licensed. Another approach would be simply to come in under the umbrella of clinical psychology. This has its disadvantage in that many of us would not want to learn things that we do not value, and we would not generally consider them qualified to teach us some of the things we do value. Many of us find the licensing alternative equally repugnant because we would have to spend a lot of time telling people what they have to learn and what they have to teach others. The alternative of doing nothing is not entirely satisfactory, either.

●ROBERT HAWKINS:

I am very averse, as I suspect a great many of you are, to the notion of licensing. It seems to me that it would produce a static profession. I would have to opt for doing nothing rather than licensing. At the same time, however, it seems that it might be functional to write some general guidelines. A collection of behaviorally oriented people from various settings might write some general prescriptions that could be circulated among state administrators, local administrators, and the sort of people who are going to hire behavior modifiers. Perhaps it would be sufficient to give them some clues about things to look for, what training usually consists of and how to find out whether someone is competent.

●WILLARD DAY:

I would not have thought that it was the natural thing for a behaviorist to want to set up licensing as an important part of the overall contingencies. If we are going to change something, it seems to me that we should ask what sort of overall changes we can make which will enable us to fit in with natural social contingencies. The idea occurs that if you have to license someone to be a behavior modifier, then eventually you are going to have to license people to be people.

●HOWARD SLOANE:

I tend to agree with the last two speakers; however, I would like to expand on the idea of changing social contingencies. It would seem to me that before we become concerned with things like licensing or setting up criteria for training, we should know something about what we are licensing and what kinds of things we want to train. I do not think we know that now. I do not think we have any demonstrated, highly effective training programs that we can say are models. We do not have anything solid we can say about what everybody should learn. Right now, I think a move in that direction would be premature and only inhibit change and development. Specifying competences in a licensing situation really concerns me. For example, when I think of the kinds of things we are doing now compared with our professional activities two years ago or four years ago, things are really changing too quickly for the kind of specification that is implied in licensing procedures. I really feel that licensing and general criteria depend upon having enough experience to feel confident in defining your product. I don't feel that confident at present.

●ROBERT HAWKINS:

Perhaps the sentiment is great enough among competent behavior modifiers AGAINST licensing that what we should do is write guidelines suggesting specifically that administrators NOT call for licensing. For a number of reasons licensing may be an attractive step to some administrators. If we don't do something like this, a licensing movement may get started and end up being something that we don't want at all. Maybe we should take a specific stand against licensing and make very public our reasons for opposing it.

●JACK MICHAEL:

One of the initial steps many professions make is to form a professional organization, and prior to any licensing, set up committees that try to determine guidelines. Perhaps, we are at the stage where some kind of general meetings where opposition to licensing and other approaches to the issue are considered. The general consensus seems to be that licensing as in electrical engineering, clinical psychology and medicine, is very unsatisfactory and that most people would be opposed to it.

•TODD RISLEY:

There is another piece of information which has not arisen here. There is currently a move to establish a certification board for "behavior therapy." The plan is to have an appointed group of noted, semi-noted and (to me) unknown people constitute a board which will pass judgment on subsequent people's membership status. A number of people have expressed opposition to this approach.

In considering certification, the proposed APA Guidelines on human experimentation and the 300 or so articles submitted to *Journal of Applied Behavior Analysis* in the past year, I have come to the conclusion that we should not license people, but that we should license procedures. I think we can formally classify procedures as those which are established and documented and those which are still experimental. Perhaps we can establish a committee which reviews procedures periodically and determines when a procedure is no longer experimental, but should be considered a standard therapeutic procedure. Anything that is not in that category should be considered experimental and all of the checks and balances on human experimentation should be applied. My tentative solution, therefore, is not to license people but to certify, specify, or define and delineate standard procedures which are no longer of interest to the experimenter but which should be readily applied as therapeutic.

•AUDIENCE MEMBER: (Beth Sulzer-Azaroff)

I think that one of the first things we should do is define "behavior modification." I have read several different definitions and I don't think there are any acceptable standards. For example, would you call a child who is a tutor a behavior modifier? I prefer to call him a contingency manager. We can talk about someone who is an aide in a hospital or a teacher or a college student as a behavior modifier, and yet I would prefer not to call him a behavior modifier, because he is not bound by the constrictions that sophisticated behavior modification requires of itself. We should restrict the name "behavior modification" to an analysis of behavior that is controlled, objective and measurable. If we adopt such a definition, we no longer have the dilemma that the clinical psychologist has because we have data that demonstrate accountability within our profession. Anyone calling himself a behavior modifier must be able to present evidence that he has been able to function in the role of collecting objective data, designing behavioral procedures, and hopefully bound by some kind of ethical restrictions.

●JON BAILEY:

I want to make a comment about the consumer pressure for professionalism. It seems to me that we have caused part of the pressure ourselves. I have yet to know a behavior modifier who was not also trying to "sell" a system. He wants to "sell" the approach to anyone who will buy it. As a result we have made statements to the effect that, "Well, anybody can do this. You don't have to be a Ph.D. All you have to know are a few basic principles and you can change behavior." We were teaching kids to modify behaviors of other kids. And, to some extent we are going to worsen the problem by creating demands that we cannot meet. If the demand for trained behavior modifiers exceeds what we can supply, then we are going to attract people on the fringes who took one course in learning and trained an animal to press a bar. Then somebody says, "Hey, I've got a kid with a problem" and they take it on. The reason they hire the guy is simple--they do not have somebody who really knows what he is doing. Perhaps we should do a little more training and a little less selling.

●ROBERT HAWKINS:

I don't know whether this summarizes it, but it seems that there are two major issues. First, there is a need to protect our scientific status as researchers so that we are not told by others what kinds of things we can investigate. That is, we want to continue as scientists without some outside agency telling us whether we are proceeding adequately or not. On the other hand, I think the field has grown large enough so that we must be accountable for what we do, not in our laboratories, but in our applied work. In some sense, we should protect the public from people who claim to be behavior modifiers but who aren't. The term "behavior modifier" is not a set of credentials. It's a term we use for people we know.

●DAVID BORN:

I would like to make a comment. There is another move softer than licensing that might solve a couple of the problems, and that is to accredit behavior modification training programs. Presumably APA's Division 25 could select a committee that would review programs at the university level and accredit programs that, by whatever their standards, were acceptable. The advantage of this is that consumers who are interested in hiring behavior modifiers would have a specific place to go and furthermore, departments would have some way of knowing whether or not they had a reasonable behavior modification training program.

●JACK MICHAEL:

This is the same step clinical psychology took even before it urged licensing. They urged standardized training programs and sort of licensed departments to offer the degree. There are obviously many solutions, but none of them are entirely satisfactory. Although many of them seem quite repugnant, non-solution may also be undesirable.

●AUDIENCE MEMBER: (Barney Salzberg)

I am a little concerned. I am recently out of school and I am not in a university setting. Frequently, I come in contact with people who do behavior modification, or claim to be doing it. On more than one occasion, I've been a little concerned about their lack of procedure. I do believe that you have to be concerned for the consumer--to deny that fact makes no sense. People who have never had a course in contingency management claim to be behavior modifiers. I think that's fraudulent. Although licensing may not be the way to go, I think we must come up with some way to tell the public how to discriminate between a genuine behavior modifier and someone who only says he is. It is one thing to disseminate information for somebody to use to effect changes in their own lives. But, when you expect a person to effect changes in somebody else's life, you have encountered a responsibility you cannot avoid. I think it is time for behavior modification to consider whether licensing _is_ the final answer. It may not be, but we should have some very firm way of saying, "Yeah, this guy meets ten out of the twelve things we think good behavior modifiers should do, or training they should have had, or experiences that they have undergone."

●ROBERT HAWKINS:

We are running short of time, but I would like to suggest that we do this kind of thing again in the future.

Training Behavior Modifiers: A Summary and a Proposal

•DON R. THOMAS

There are a number of current issues in behavior modification which we can no longer afford to ignore. Recently, *Psychology Today* (November, 1972) published an article describing many of the activities and programs of behavior modifiers. As the field continues to grow, we can no longer dismiss lightly mundane issues such as standardized training programs and licensing. The days when it was possible to be on a first name basis with all of the behavior modifiers in the country no longer exist. No longer is it possible to list all of the behavior modifiers by listing the past students of Skinner, Bijou, Baer, Michael, Azrin and the rest. No longer is it possible to list behavior modifiers by reference to the university from which they were graduated--most of the programs which produced behavior modifiers in the 1960's no longer exist, because their staffs have moved to other places.

The panel discussion on training behavior modifiers did not produce a consensus of opinion on the next appropriate step; however, it clearly raised some important issues. First, there has been a recent increase in the number of documented instances in which incompetent individuals have claimed to be behavior modifiers and have conducted programs in the name of behavior modification. Second, there is a general aversion to the notion of licensing. Rob Hawkins best expressed the resulting dilemma when he pointed out that although none of us want to be told how to act or how to collect data, all of us feel a need to protect ourselves and the public from people with no training who claim to be behavior modifiers.

A variety of partial solutions was recommended. Training programs in behavior modification might be accredited. General procedures which have been empirically demonstrated to be effective might be certified. A professional organization might be formed to generate guidelines. Division 25 of the American Psychology Association might begin to accredit training programs. Finally, it was suggested that we ought to continue to talk about the issue. Clearly, we can continue to talk about the issue indefinitely.

I propose that time be allocated during the 1973 Kansas Conference on Behavior Analysis in Education to continue the discussion of issues raised last Spring. But more important, I propose that the objectives of the discussion be clearly defined as the development of a professional organization whose members are prepared to search actively for solutions to the problem. The need to protect ourselves and the public is real. As Jack Michael asked, "Will the REAL behavior modifier please stand up?"

⁂

PART 2:

current topics in BEHAVIOR ANALYSIS

BEHAVIOR ANALYSIS OF DRUG EFFECTS IN THE CLASSROOM

stephen i. sulzbacher[1]

The use of pharmacologic agents to modify the behavior of children in elementary school classrooms is currently causing considerable concern and controversy among physicians and educators. Although drugs have been used as a treatment for learning and behavior disorders for more than thirty years, very little is known about the exact nature of the effects of these drugs on behavior, due largely to the inadequate methodology which has been employed in studies of drug effects (Sprague and Werry, 1971; Hollis and St. Omer, 1972).

Behavior analysis methodology has been systematically applied to the study of drug effects in animals (Sidman, 1956; Thompson and Schuster, 1968), but it has rarely been applied to the analysis of drug effects on educationally relevant behaviors of children in the classroom. This study presents representative cases illustrating the effects of behavior modifying drugs on academic performance and classroom behavior of elementary school children and presents a methodology for the objective assessment of drug-behavior interactions.

Review of Previous Research[2]

A literature review was conducted on all studies employing psychotropic medication with children which were published between January, 1937, and March, 1971. These studies were categorized on the basis of methodology employed and according to the nature of the dependent variables studied. A summary of these studies is presented in Table 1, which shows that only a very small percentage of the available research employed an objective measurement of behavior. It can be seen from the data presented in Table 2 that the likelihood of obtaining statistically significant results was not related to the type of experimental design employed, but was strongly influenced by the response measurement technique employed. Table 2 reveals a disturbing inverse relationship between degree of rigorousness of response measurement and

---

[1] Department of Pediatrics and Child Development and Mental Retardation Center, University of Washington. Research supported by Maternal and Child Health Service Project #913.

[2] This literature review is adapted from an article which will appear in *Pediatrics* and should not be reprinted without permission.

TABLE I

SUMMARY OF SURVEYED LITERATURE (1937-1971)

TOTAL NUMBER STUDIES SURVEYED.	1,359
STUDIES OF UNRELATED *(non-behavioral)* EFFECTS OF PSYCHOACTIVE DRUGS.	- 560
UNAVAILABLE REPORTS.	- 43
TOTAL NUMBER STUDIES OF DRUG EFFECTS ON THE BEHAVIOR OF CHILDREN.	756

	NUMBER	PERCENTAGE OF TOTAL (n = 756)
UNCONTROLLED STUDIES *(no placebos or double-blind)*.	548	72.5
CONTROLLED STUDIES CATEGORIZED BY DEPENDENT VARIABLE:		
a) *clinical impression* or professional opinion.	34	4.5
b) rating scale	99	13.1
c) standard psychological tests.	46	6.1
d) direct measurement of behavior	29	3.8
	756	100.0

the finding of significant differences in results. Significant differences indicating a favorable drug effect are found in an overwhelming majority of the studies employing "professional opinion" or rating scales as the dependent variable, but such results are considerably less likely to be found when standardized psychological tests or direct measurements of behavior are employed.

As will be seen in the following study, drugs do in fact exert powerful effects on human behavior, some of which are quite beneficial; but, these vary greatly among individuals and it will be shown that continuous and direct measurement of educationally relevant behaviors in individual subjects provides more meaningful data on drug effects in children than is currently available with other measures.

TABLE 2

ANALYSIS OF STUDIES BY TYPE OF DESIGN, MEASUREMENT TECHNIQUE AND SIGNIFICANCE OF RESULT

RESPONSE MEASUREMENT TECHNIQUE EMPLOYED	SIGNIFICANT DIFFERENCE FAVORING THE DRUG		RESULTS NOT SIGNIFICANT OR FAVORING PLACEBO	
	Within-Subjects Design	*Between-Groups Design*	*Within-Subjects Design*	*Between-Groups Design*
DIRECT MEASUREMENT OF BEHAVIOR	8	4	10	7
STANDARD PSYCHOLOGICAL TESTS	2	6	12	26
RATING SCALE	19	37	17	26
"CLINICAL IMPRESSION" OR PROFESSIONAL OPINION	13	17	4	0
TOTALS	42	64	43	59

x^2 TEST OF ASSOCIATION BETWEEN SIGNIFICANCE OF RESULTS AND MEASUREMENT TECHNIQUE EMPLOYED

	SIGNIFICANT DIFFERENCE (a)	NO SIGNIFICANT DIFFERENCE	(n)
DIRECT MEASURE	12	17	29
PSYCHOLOGICAL TEST	8	38	46
RATING SCALE	56	43	99
OPINION	30	4	34
	A = 106	102	N = 208

$$x^2 = \frac{\Sigma\frac{a^2}{n} - \frac{A^2}{N}}{k(1-k)}$$

$$k = \frac{A}{N}$$

$$x^2 = 42.716$$

$df = 3$; $p < 0.000001$

METHOD

Subjects

Data from three subjects will be reported. Ralph was seven years, eleven months at the beginning of the drug study, had a full-scale Wechsler Intelligence Scale for Children (WISC) intelligence quotient of 81, and had been previously diagnosed as autistic and minimally brain damaged. Joe was eight years, five months old at the beginning of the study, had achieved a full-scale WISC I.Q. of 60 when tested without medication prior to the study, and a score of 71 on the same test on medication after the study. He had previously been diagnosed as emotionally disturbed, dyslexic, and minimally brain damaged. Gordon was seven years, eight months old at the beginning of the study, had a full-scale WISC I.Q. of 104 and had been diagnosed as hyperactive and learning disabled.

All three boys came from fairly stable homes and attended Special Education classes in public schools in the Seattle area. The drug studies with Ralph and Joe were conducted while they attended a special summer classroom as part of a research project at the University. The study with Gordon was done in a public school class for children with learning disabilities.

Because of the imprecision of the diagnostic categories of minimal brain dysfunction and hyperactivity, the decision to treat these particular children with stimulant medication was essentially judgemental. The judgements were based on clinical opinion that the child was overactive or inattentive in the classroom setting and that he was not learning at the rate of his peers. At the present state of knowledge, the only reliable way to predict which child will respond favorably to stimulant medication is to observe his behavior on and off the drug (Sprague and Werry, 1971).

Drug Protocol

The medications employed in the present study were dextro-amphetamine sulfate (Dexedrine^r) and methylphenidate (Ritalin^r). Dosages used were between 5 mg. and 20 mg. These drugs are central nervous system (CNS) stimulants, and although there are a number of theories about how they act, the precise mode of action in humans is still unclear.

Of the three general classes of drugs currently prescribed for behavior problems--CNS stimulants, sedative-hypnotics, and anti-convulsants--the CNS stimulants have certain methodological advantages. Whereas some of the other drugs remain active in the body for at least a week, CNS stimulants are almost completely (70%) metabolized within 24 hours. The use of these drugs as *independent variables* makes it possible to manipulate dosage level on a day-to-day basis. In each case, at least two different dosage levels are used as well as a placebo condition.

Experimental Design

The basic ABA reversal design with multiple replications was employed in all studies. Drugs are uniquely suited as independent variables with this type of design since they can be imposed and withdrawn at will. Each study takes about a month (about 20 school days) yielding 5-6 days of data under each drug condition. A random number table was employed to assign one of the three drug conditions to each of the 20 school days. An example of a random reversal drug schedule is shown in Table 3.

TABLE 3: *RANDOM DRUG SCHEDULE*

	PLACEBO	5 MG. DEXEDRINE^{R}	10 MG. DEXEDRINE^{R}
FEBRUARY 12		X	
FEBRUARY 13	X		
FEBRUARY 14[a]		X	
FEBRUARY 15[a]		X	
FEBRUARY 16		X	
FEBRUARY 17	X		
FEBRUARY 18			X
FEBRUARY 19			X
FEBRUARY 20			X
FEBRUARY 21[a]	X		
FEBRUARY 22[a]		X	
FEBRUARY 23			X
FEBRUARY 24	X		
FEBRUARY 25	X		
FEBRUARY 26			X
FEBRUARY 27		X	
FEBRUARY 28[a]		X	
MARCH 1[a]	X		
MARCH 2			X
MARCH 3		X	
MARCH 4		X	
MARCH 5			X
MARCH 6			X
MARCH 7[a]			X
MARCH 8[a]			X
MARCH 9	X		
MARCH 10		X	
MARCH 11	X		
MARCH 12		X	
MARCH 13	X		
TOTAL NO. OF DAYS. .	9	11	10
TOTAL NO. OF SCHOOL DAYS.	7	7	8

---

[a] Weekend days.

A triple-blind procedure was employed in which neither the subjects, the prescribing physician nor anyone taking data was aware of the dosage level being administered on any given day. The medication was given to the parents in small envelopes, each dated and containing one day's medication. The same number of tablets was in each envelope. The parents were unable to distinguish drug tablets from placebos, since the placebos supplied by the drug manufacturers (Ciba-Geigy Corporation and Smith, Kline and French Laboratories) were identical in appearance with the active tablets. Each child was given the contents of his envelope in the morning before coming to school and the parents were instructed to date and return the empty envelopes at the end of the study.

Classroom Conditions

The experimental classroom was about one fourth the size of classrooms customarily seen in the public schools, but was equipped with a one-way observation window and microphones. The adjoining observation booth was used for data-gathering and to enable parents to watch the classroom procedures. Another small room adjoining the classroom was used for individual tutoring sessions.

The class met four days a week for six weeks, each session lasting from 9 a.m. until noon. The total class enrollment was six, two of whom were included in the present drug evaluation study. When a child arrived at his desk in the morning, he found a file folder containing his daily assignments and a sheet of paper listing exactly what work he had to complete that day. The children were told that all the work in the daily folder had to be completed correctly before they would be permitted to leave. The school work was individually prescribed according to the child's academic needs. Although there were only one or two teachers in the classroom at any given time, a total of five teachers were involved in the study, and this team carefully analyzed each child's daily work so that assignments were based on a careful analysis of performance and errors on his previous work. All of the academic work was either with individually prepared teaching programs or commercially available programmed textbooks. All of the work was timed and rate per minute of errors and correct responses were computed for each subject each day.

The children were told that when they completed their work they would be allowed free time to play outside or with toys available in an adjoining playroom. They also received marks in a small *point-book* for staying in their seats and not talking during class time. These marks were redeemable for money or special treats and the boys were also allowed to earn extra marks by doing extra schoolwork instead of playing when they had completed their daily assignments.

From the child's point of view, the classroom management procedures were quite simple: he was required to complete a fixed amount of work each day and correct all of his errors. When this work was completed, he would be allowed to play or go home. Each child knew he would be taken from the classroom daily to participate in other experiments in another room and could earn marks during these sessions.

The case of Gordon is presented to demonstrate that comparable data to that obtained in the experimental classroom situation could be gathered by a public school teacher under the conditions prevailing in a Special Education class. Gordon's case is included in the present report to illustrate that meaningful drug dosage decisions can be made on the basis of public school classroom data, even though one must be satisfied with a smaller quantity of information gathered under less clearly defined conditions.

Reliability

No reliability scores were computed for any of these data, since there was only one person observing at a time in the experimental classroom and since the public school teacher recorded all her own data. The principal investigator reviewed all the programmed academic materials which were scored as part of the study and found no points of disagreement with the scoring of correctness of the individual responses.

Prior to each drug study, the principal investigator made joint training observations of the target behaviors of the individual children with the teachers who would subsequently record the data. In the case of each observer, a criterion agreement of 90% was reached after two 20-minute training periods. The data from these observations were, of course, not used in the study.

Whereas the absolute reliability of these data cannot be demonstrated or even inferred, several factors provide some support for the assumption that there was no systematic bias relative to the drug conditions:

1▶ Since drug conditions were randomly assigned to successive days, any change in accuracy of recording over time would affect all drug treatment conditions equally.

2▶ All the teachers observed for equivalent amounts of time each day to insure against the possibility that a given teacher's observations would fall during any one treatment condition more than any other.

3▶ All observers were unaware of the drug schedule and did not know which child had drug or placebo any given day; since they did not know whether a child had received active or placebo medication, any errors of observation were assumed to occur with the same likelihood under the different treatment conditions.

Dependent Variables

A multiple baseline procedure was used, with measurement of behavioral and academic target behaviors. It was important that the behaviors being measured with each child include both acceleration and deceleration targets, since a decrease in both of these classes of measured response would indicate sedation (overdose), an undesired side effect.

Correct and error rates were obtained using programmed arithmetic texts. Reading and writing performances were from carefully prepared comparable daily tasks in workbooks or oral recitation.

Talking out in class was recorded for all three subjects of the present study. In addition, rate of getting out of seat during class time was computed from data taken in the observation booth in the experimental classroom on the two boys who were studied there. Out-of-seat was defined as air between the chair seat and the child's buttocks or leg (the child could be in seat with a leg folded under him). Each occurrence of out-of-seat was terminated only when he returned to his chair. A talk-out was any vocalization which wasn't solicited by the teacher. This definition included grunts, humming, and audible talking to self. An incident of talking out was terminated by someone else interjecting a comment to the child or by a silent pause of two or more seconds.

RESULTS

Case Study I

Correct and error rates were kept on Ralph's performances in three academic subjects. His daily arithmetic assignments were from the Rainier Math Program (Edmark, 1972). All of the assignments were computation facts involving no numbers with more than two digits and included addition with carrying, and subtraction without borrowing problems. Oral reading rate was obtained from *Around the Corner* (Ginn, 1966). After he completed the reading task, Ralph had to make written responses to comprehension questions about the story he had just read.

Ralph was exposed to three drug conditions: *placebo, 5 mg.* and *10 mg. dextroamphetamine*. The following data are means of nine placebo days, seven 5 mg. days and six 10 mg. days, which occurred in random order during the study. His frequencies of talking out and of getting out of his seat during class are plotted for each drug condition in Figure 1.

Both of these behaviors showed a marked reduction on days when active medication had been administered. Chi-Square analyses confirmed the statistical significance of these differences ($p < 0.01$).

Drug effects on Ralph's academic performances are displayed in Figure 2. His arithmetic performance rate was more than doubled under the drug conditions, while errors remained unaffected. However, both language arts behaviors showed a non-linear dose-effect relationship, with performance rates under 10 mg. being lower than in the placebo condition.

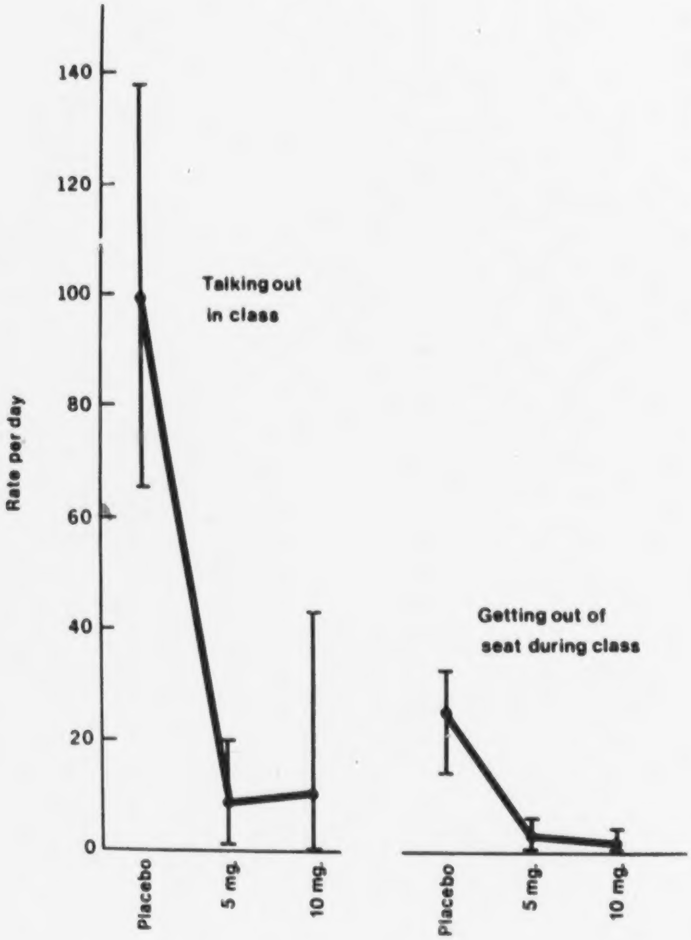


Fig. 1. Effects of d-amphetamine on Ralph's classroom behavior. The heavy lines connect the means of each condition. The light bars around each mean indicate the range of daily rates.

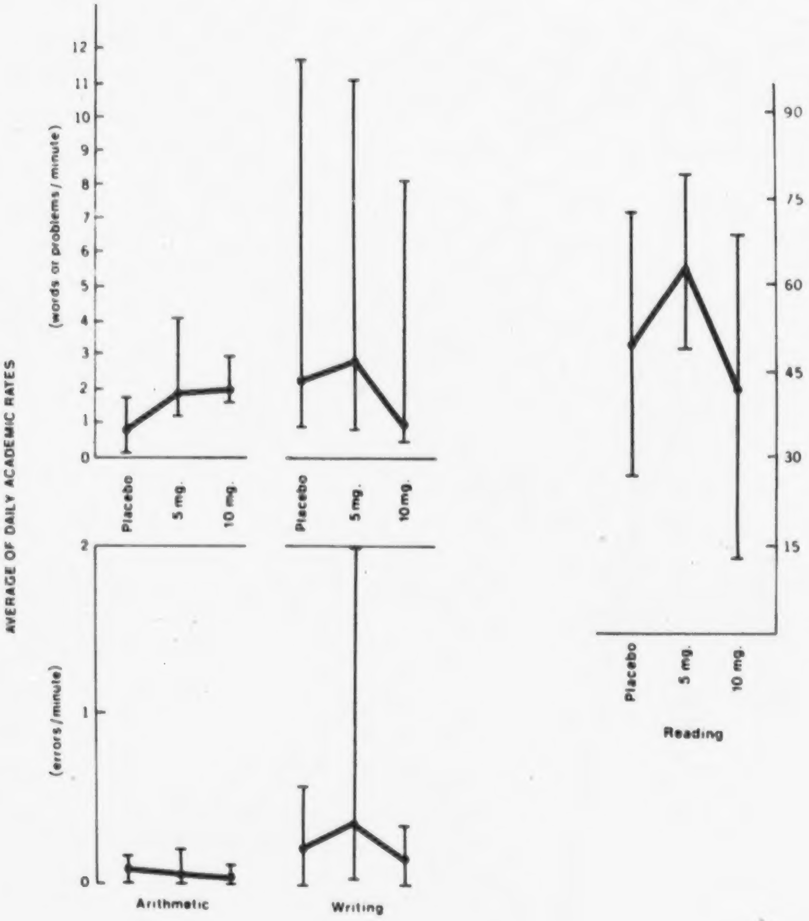


Fig. 2. Effects of d-amphetamine on the academic performance of Ralph.

Case Study II

Joe's curriculum consisted of written number chaining and beginning addition facts from the Rainier Math Program, assignments from a programmed writing text, *Handwriting with Write and See* (Skinner and Krakower, 1968), and a beginning phonics reading program devised by the present investigator which consisted of discrimination tasks between vowels, consonants, and phonetic letter combinations.

There were seven school days on which Joe received 10 mg. of dextroamphetamine, six days on which he received 5 mg. and six days on which he received placebos. The unequal number of days under the three conditions was due to Joe's absence from school on two days. The mean frequencies of two disruptive behaviors under each drug condition are shown in Figure 3.

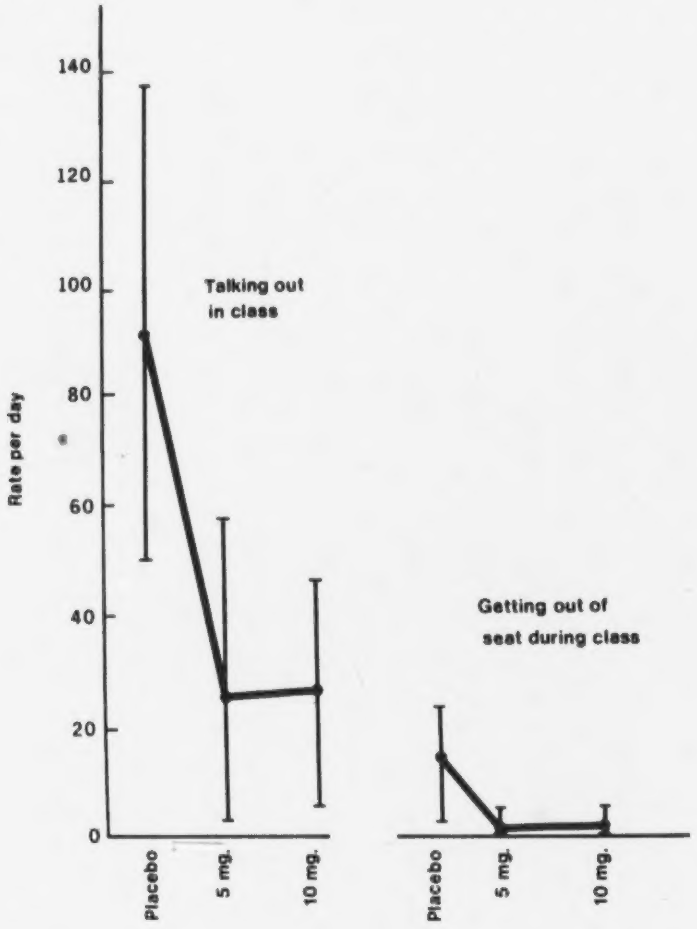


Fig. 3. Effects of d-amphetamine on Joe's classroom behavior.

There were highly significant decrements of both talking out and getting out of seat on the days Joe received active medication ($p < 0.01$). The effects of medication on Joe's academic performance are shown in Figure 4. The total performance rate in arithmetic and reading is slightly increased under the 5 mg. condition; the performance increase can be seen to result almost entirely from increases in the error rate. Joe's writing rate, however, shows an increment at 5 mg. and a further increment at 10 mg., with little concomitant increase in error rate.

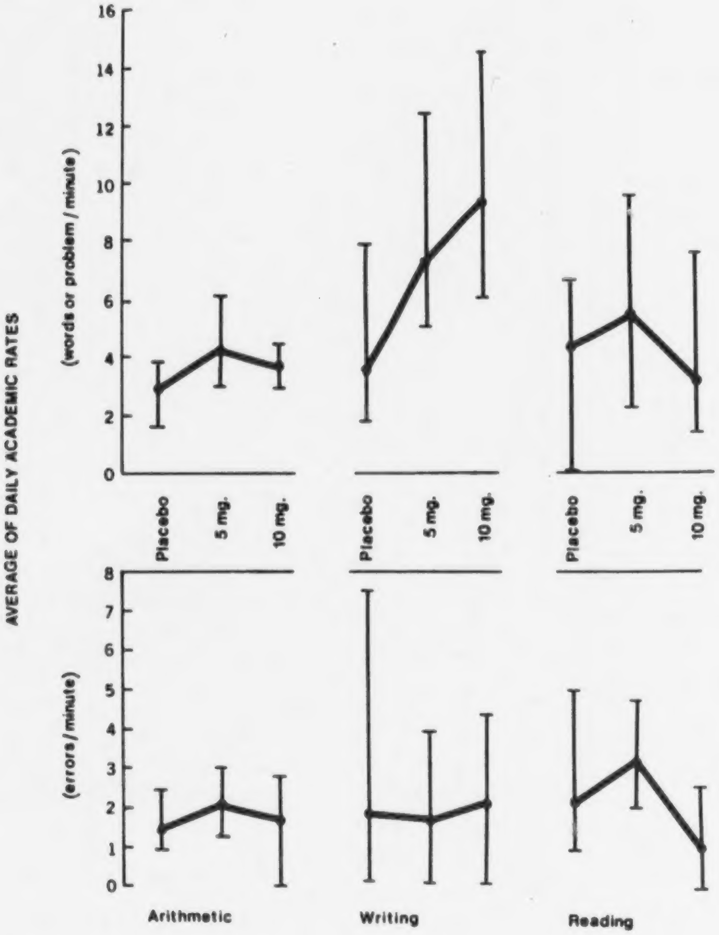


Fig. 4. Joe's academic performance rates under placebo and two dosages of d-amphetamine.

Case Study III

Data were obtained on Gordon's behavior for twenty school days in his public school classroom. There were five days of baseline, four days of placebo, six days with 5 mg. of methylphenidate and five days with 15 mg. The teacher in Gordon's class for children with learning disabilities gathered data on reading and on talking out in class. The reading tasks consisted of daily assignments in the workbook which accompanies *On Cherry Street* (Ginn, 1966). Each incident of speaking or vocalizing without being asked to was recorded by the teacher on a golf-stroke counter which she wore on her wrist.

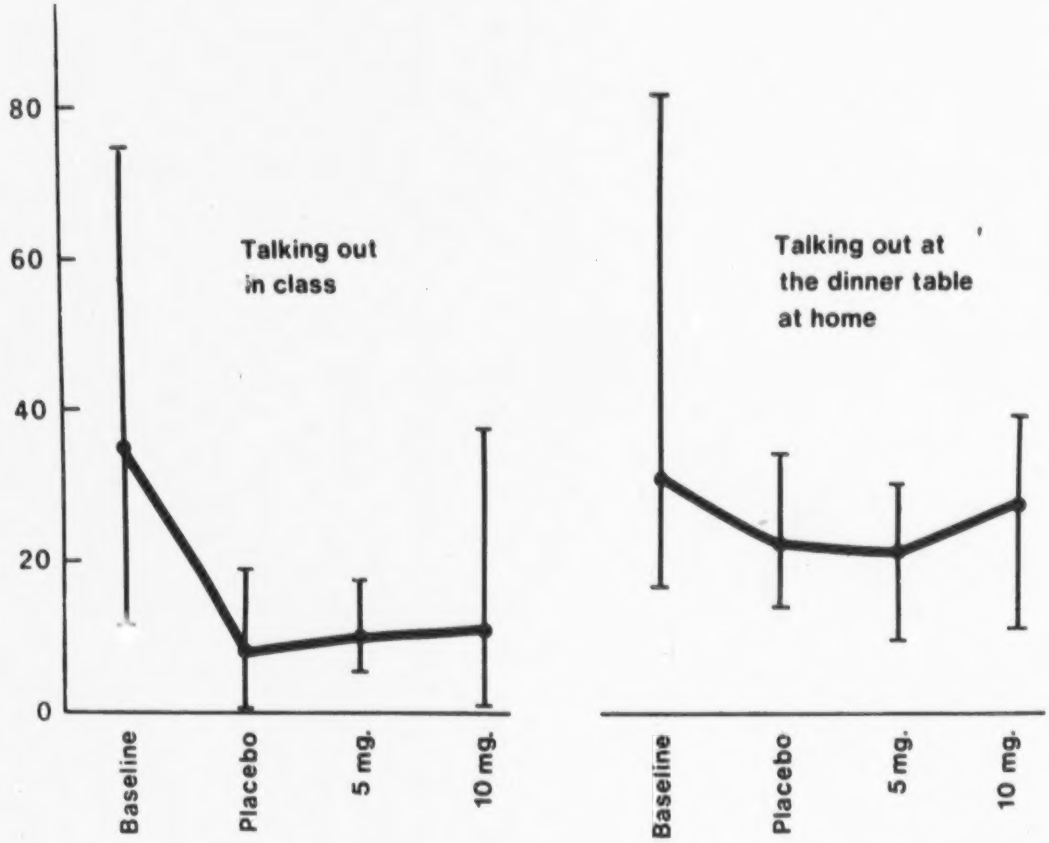


Fig. 5. Effects of methylphenidate on Gordon's talking behavior.

The mother's complaint at home was that Gordon frequently interrupted other people in the household and that this was particularly prevalent and annoying during dinnertime conversations. The mother recorded all incidents of talking, whether appropriate or not, at the dinner table and also noted the time dinner began and ended so that rates could be computed.

A summary of Gordon's responses under the various drug conditions is shown in Figure 5. A fairly clear-cut change in Gordon's talk-out rate is apparent between baseline and all other conditions ($p < 0.01$), suggesting a placebo effect rather than a change due to the actual drug.

Gordon's rate of talking at the dinner table at home was apparently unaffected by the medication. In Figure 6, his language arts performance rate does not appear to show any systematic effect of medication, whereas the error rate showed a non-significant placebo effect. With additional reduction in error rate as successively higher drug dosages.

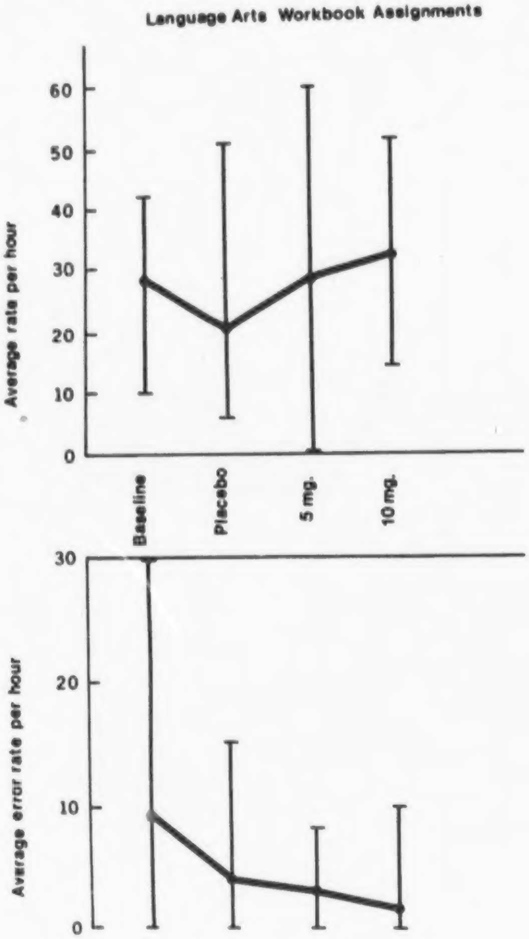


Fig. 6. Effects of methylphenidate on Gordon's language arts performance.

SUMMARY

Three cases were presented which systematically replicated procedures for assessing, in public school classrooms, the effects of psychotropic drugs on learning and behavior. In each case, different behaviors were shown to be affected by the medication, illustrating the degree of individual differences in drug response.

The assessment of drug effects in the classroom appears to be a problem which, due to our present state of knowledge of pharmacology, can only be solved through the use of single subject studies of individual drug responses on the educationally relevant behaviors which led to the child's being referred for treatment. The cases presented indicate that central nervous system stimulant medication can effectively modify disruptive behaviors without adversely affecting academic performance in the classroom. This is not to suggest, however, that other behavior modification procedures would not be equally effective in some cases.

REFERENCES

Hollis, J. H. and St. Omer, V. V. Direct measurement of psychopharmacologic response: effects of chlorpromazine on retardate behavior. *American Journal of Mental Deficiency*, 1972, 76, 397-407.

Sidman, M. Drug-behavior interaction. *Annals of the New York Academy of Sciences*, 1956, 65, 282-302.

Sprague, R. L. and Werry, J. S. Methodology of psychopharmacological studies with the retarded. In N. R. Ellis (Ed.), *International Review of Research in Mental Retardation*, 5, New York: Academic Press, 1971, 147-219.

Thompson, T. and Schuster, C. R. *Behavioral Pharmacology*. New York: Prentice-Hall, 1968.

THE EFFECT OF GENERAL & DESCRIPTIVE REINFORCEMENT ON "CREATIVITY" IN EASEL PAINTING[1]

elizabeth m. goetz[2]
mary m. salmonson

Early childhood educators have traditionally paid lip service to the development of creativity even though it is their appropriate concern (Evans, 1971). Perhaps research has not concretely presented to educators methods for the cultivation of creativity.

Creativity is a difficult concept to define objectively. Definitions of the concept have been used in reference to such diverse areas as the mental processes and abilities, personality, products, verbal behavior and physical behavior (Torrance, 1965). Only the products and verbal and physical behaviors lend themselves to objective measurement in that they may be analyzed as to the specific novel forms they contain. A key criterion for creativity is originality, difference or novelty.

Goetz and Baer (1971) indicated that it was possible to increase diversity of blockbuilding in preschool children by reinforcing the use of new block forms for each child. The present study continued the investigation into this aspect of creativity by using a new medium--easel painting, and also studied specific types of teacher attention that may be related to an increase in creative behavior.

METHOD

Subjects

Three preschool children from the Edna A. Hill Child Development Preschool Laboratories were selected for this study. All three were selected due to a fairly low number of different forms used in their easel paintings.

---

[1] This research was supported in part by P.H.S. research grant HD02528 National Institute of Child Health and Human Development.

[2] Requests for reprints should be sent to Elizabeth M. Goetz, Edna A. Hill Child Development Preschool Laboratories, Department of Human Development, University of Kansas, Lawrence, Kansas 66044.

Procedure

The subjects (Sue, Carrie, and Jane--pseudonyms for purposes of reporting) were invited to paint at the preschool easel by the teacher-experimenter. Three different color cups (red, blue, and yellow) with a 1/2" brush in each were always available. The same size paper (18" x 20") was always used. The children determined the length of each easel painting session by stopping when they felt their pictures were completed, and at this time the experimenter thanked the subject for painting.

The experimenter defined 25 painting forms from many children's pictures prior to the beginning of the study (Table 1). Some of the forms were curve, mass, spatter, zig-zag, and triangle. Reliability on these forms was found to be possible.

Teacher attention delivered contingently on each new form as it was painted was defined by two categories:

1. GENERAL REINFORCEMENT - A painting form was acknowledged as being "good" but no specific aspect was singled out for comment. For example, general reinforcement took the form of such phrases as, "Isn't that nice," or "See what you have here," when the child painted different forms.

2. DESCRIPTIVE REINFORCEMENT - The teacher's comments were directed toward a specific form describing it. Such comments as, "That is a very straight horizontal line you have painted," or "Now you are doing a zig-zag back and forth," are examples of descriptive reinforcement.

A third condition of no teacher attention was also used. The child simply painted while the teacher watched, but made no comments on the child's painting.

Each child was run in an individual analysis design and therefore served as her own control across various conditions. In addition, each child was run in a different sequence of conditions so that the effects of each condition in a different order could be studied. The design for Sue was: *ABCBC*; for Carrie: *ABCAC*; and for Jane: *ACAC* (where *A = No Reinforcement; B = General Reinforcement;* and *C = Descriptive Reinforcement*).

Reliability checks were taken by an independent observer on painting forms, teacher attention and duration of painting during each condition for each subject.

TABLE I

DEFINITIONS OF PAINTING FORMS

The horizontal, vertical, and diagonal lines are defined by angles relative to the bottom of the paper, using the smallest angle as a measurement.

1. BLENDED COLOR - any hue formed by mixing two or more pure or available colors onto the paper.

☆2. CIRCULAR - any nearly enclosed curve such as circles, ovals, ellipses, etc.

3. CURVE - a line or any part of a line, at least 3" long, continuously bent so that no portion of it is straight.

4. DIAGONAL LINE - a relatively straight line, at least 3" long, forming a 10° - 80° angle.

5. DUPLICATE FORM - a relatively exact pair of any abstract form, clearly seen as a design, or any of the starred forms: The size and color may vary but not the structure. It should be basically the same. Simple forms, such as circles, require more exactness than more complex forms, such as an irregular enclosure or a simulation. A staccato grouping itself is not a duplicate. The same grouping must be repeated in another area of the paper.

6. HORIZONTAL LINE - a relatively straight line, at least 3" long, forming a 0° - 10° angle.

☆7. IRREGULAR ENCLOSURE - any closed or nearly closed *unsymmetrical* line formation.

☆8. LAYERED - three or more repeated lines, using two or more different colors, one inch or less in width, which lie side by side. Each line should be a different color than the one(s) directly beside it.

9. MASS - any nearly solid painted area at least 2" long and 2" wide.

10. OVERLAPPING SAME FORMS - a duplicate with one form extending over and covering a part of the other.

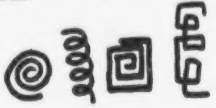
11. PATTERN - same as a duplicate, only requires three or more copies.

☆12. RECTANGULAR - any nearly enclosed form with four relatively straight sides and four 90° angles, approximate within 10°. Opposite sides must be relatively parallel.

☆13. SIMULATION - a construction which resembles a real-life object. Symbols, such as letters and numbers, are not simulations.

14. SPATTER - three or more scattered or dashed small particles or drops resulting from one swing of the brush.

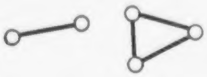
15. SPIRAL - a winding or coiled line; whether consisting of a continuous curve or straight lines. There must be at least two complete revolutions.



16. STACCATO - three or more quick dabs that are not the result of a spatter and are clustered in the same area. They may or may not be overlapping.

☆17. SYMBOL - something other than a simulation which stands for something else, such as signs, letters and numbers.

☆18. TINKER TOY LINE - circular forms with one or more straight lines connecting them.

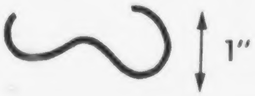


☆19. TRAIN OF COLORS - a series (three or more) of two or more colors forming a line or procession.

☆20. TRIANGULAR - any closed or nearly enclosed form with only three sides and three angles. The INSIDE lengths of at least two (2) sides must be at least 1 1/2".

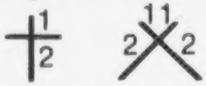
21. TRICKLE - a flowing or falling drop in a small, broken or gentle stream.

22. UNDULATING LINE - a line with three or more curves approximately 1" or greater in depth.

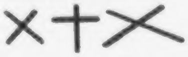


23. VERTICAL LINE - a relatively straight line, at least 3" long, forming an 80° - 100° angle.

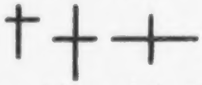
24. X-SHAPE OR CROSS - two interesting straight lines of linear proportions not greater than *two-to-one* (2:1). The two portion of an intersected line must be in proportions no greater than *two-to-one* (2:1), also. Examples:



   1. If the lines are of relatively EQUAL LENGTH, the angle of intersection is arbitrary, but the lines must intersect at relatively the same point on each line. Examples:



   2. If the lines are not of equal length, the angle of intersection must be relatively close to 90° (+ or - 10°). Examples:



25. ZIG-ZAG LINE - a line or any part of a line with three or more angles formed by turning first to one side and then to the other.

RESULTS

The results of the three subjects were similar under the same conditions regardless of the order they appeared. It is important to note that the painting times which ranged from approximately four minutes to under one minute had a slight tendency to become shorter across the twenty (or twenty-three) sessions of the study for all subjects (Figure 1).

Sue (top graph of Figure 2) had an average of 7.67 different forms under no reinforcement (baseline) condition across the three sessions. When general reinforcement was started on the fourth session, the average number of forms did not increase over baseline. However, the two conditions of descriptive teacher attention (interspersed with another general teacher attention) did show an increase in number of different forms used.

Carrie (middle graph of Figure 2) also showed the highest number of different forms used when descriptive attention was used.

Jane (lower graph) showed a slight increase across baseline conditions (first and third conditions) but when descriptive reinforcement followed it was always higher than baseline (second and fourth conditions).

For all three subjects, descriptive praise resulted in an increment in form diversity. For two of the subjects, that had the comparison, descriptive praise resulted in higher form diversity scores.

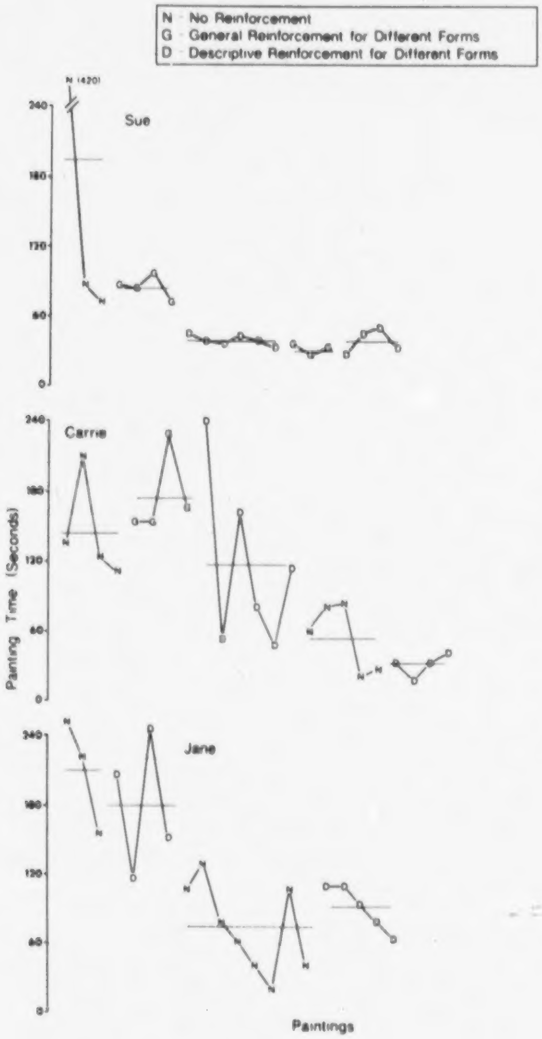


Fig. 1. The durations of time spent in each individual painting for three children in the course of easel painting training. Initial points, labeled N's, represent baseline durations under conditions of no reinforcement; points labeled as G's represent durations under conditions of general reinforcement contingent on each new form drawn in a single painting; and points labeled as D's represent durations under conditions of descriptive reinforcement contingent on each new form drawn in a single painting.

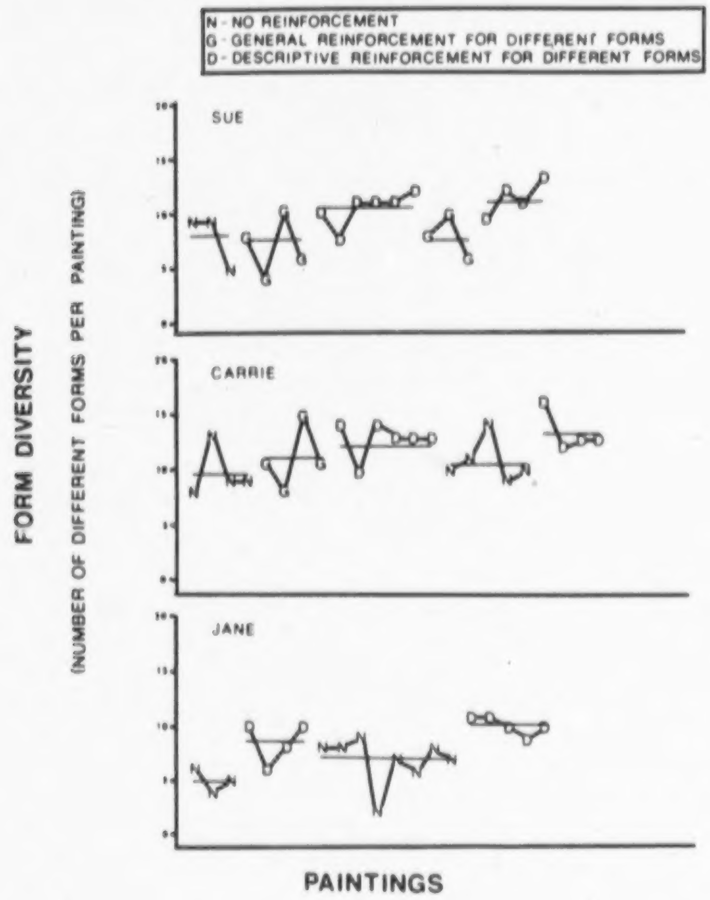


Fig. 2. The form diversity scores of three children in the course of easel painting training. Initial points labeled as N's represent baseline observations under conditions of no reinforcement; points labeled G's represent scores produced when general reinforcement was contingent on each new form drawn in a single painting; and points labeled as D's represent scores produced when descriptive reinforcement was contingent on each new form drawn in a single painting.

Reliability scores on identification of painting forms, type of teacher reinforcements used and duration of painting were ninety-nine percent or higher.

DISCUSSION

This research suggests that a teacher can, through the type of attention he gives, increase the *creative aspects* (form diversity) for children. The use of descriptive praise is the most effective and necessitates that the teacher study different forms children make so that he can identify them and respond specifically to them. Further, teachers must individually score children's easel paintings so that he may differentially respond to children's use of new forms. The results also tend to question whether or not just giving children "practice" in easel painting, without feedback, will help develop creative behaviors.

Further research is needed to differentiate the effects of general and descriptive praise not only in easel painting but in the many creative activities in which children engage.

REFERENCES

Evans, E. D. *Contemporary influences in early education.* New York: Holt, Rinehart and Winston, Inc., 1971, pp. 317-319.

Goetz, E. and Baer, D. M. Social reinforcement of "creative" block-building by young children. In Eugene A. Ramp and Bill L. Hopkins (Eds.), *A new direction for education: behavior analysis 1971.* Lawrence, Kansas: The University of Kansas Support and Development Center for Follow Through, 1971.

Torrance, E. P. *Rewarding creative behavior.* New York: Prentice-Hall, Inc., 1965.

FREEDOM & RESPONSIBILITY IN AN ELEMENTARY SCHOOL

charles l. salzberg

Freedom cannot exist long in a school unless it is accompanied by responsible behavior. This study attempted to develop and maintain responsible behavior in two groups of elementary school students regarding their mathematics program. When students were allowed to do as little or as much as they wished, they did less than one half of a page each day. When the freedom to leave class was made contingent on finishing one page of math their rate of work (problems/minute) almost doubled. A procedure in which children recorded their own progress in math on special charts generated much interest but little change in their rates of math progress. A third procedure in which students were required to pass a quiz each week in order to leave class early almost doubled the rate of work. Students in the older group were subsequently permitted to elect a condition in which they did not have to attend class at all if they passed two quizzes per week. Sixty percent of the older students elected to try this procedure. Their rate of work once again doubled. Those students were progressing at four times the normative rate of their public school-age peers. With the onset of the last two conditions students began to take initiative for their own academic progress. They began to come to class early to get a fast start. Students also began to voluntarily take some work home. They were beginning to demonstrate behavior that the teachers would call *responsible*.

The free school movement is popular and it appears to be growing. Like most movements, the free school movement is defined more by what it opposes than by those things that it favors. It grew out of a reaction against the oppressive atmosphere of traditional educational practices. It continues to grow for these same reasons, but also in part because of the progressively more public failure of these same traditional school procedures.

Free school advocates are reacting against authoritarian classrooms in which children are maximally isolated from each other in single desks, arranged in rows. They are reacting against children spending most of their day in a single, often sterile room, usually confined to one chair. There is also opposition to the behavioral priorities which exist in traditional classrooms such as: not talking, not chewing gum, not disagreeing with the teacher, not making noise. When basic skills are taught the traditional role for the student is little more than passively attending to the teacher while she runs off her chain of instructional behaviors. This is neither a very effective nor a pleasant procedure for children to endure day after day.

The goals of the free school movement then, are to open up the classroom, encourage the children to interact with each other and with stimulating materials in a rich environment so that they may learn inductively, and most important so that they may learn that the learning process itself can be a satisfying, enjoyable experience.

Offering criticism and opposition to a system is one thing. Offering practical alternatives is something else again. A large number of free schools have been started in the recent past. However, the average life span of a free school is less than two years. It is not clear why they fail. Free schools are not generally oriented toward measurement and so quantitative answers to that question are not forthcoming. Lacking that, we would like to relate some of our experiences to you.

The Colorado Springs Community School was begun in 1969 by a group of parents as an alternative to the local public school system. From the outset it was developed along the free school model. Children ranged in age from 5 to 14 years old. Classes were voluntary for the children. Teachers, for the most part, tried to help the children organize activities by providing guidance, materials, and explanations when they were requested. Children had free run of the school and the surrounding area. There were very frequent field trips to local places of interest as well as more prolonged trips for camping. Children were not coerced into participating in any activities. They were encouraged to express their thoughts and feelings. Over the course of the year academic work became less and less frequent. The best efforts of a dedicated staff with a low student-teacher ratio were not sufficient to generate very much skill oriented behavior. Even more discouraging was the development of inconsiderate, rude, and abusive social behavior of the children toward one another and toward the adults. Children showed little respect for one another's rights to privacy, to work undisturbed, to their own property, or to be free from physical abuse. On occasion group meetings were held where these issues were discussed. The children would agree that many social and property abuses were undesirable for everyone concerned. However, these verbal statements had no effect on their subsequent actions. The school had become a place where nothing and no one was safe from destruction, even the building itself. Complete freedom from adult constraints had produced an environment where none of the children were free, even from fear for their own physical well-being.

We cannot be sure about the extent to which our experience applies to other free schools. However, it would not be surprising if the above factors have something to do with why so many free schools fail in a short time.

In planning for the next year the staff and some of the parents at the Community School began to rethink the issue of freedom within the school. First, it became evident that freedom is not a global entity. In fact, there seemed to be innumerable potential freedoms within a school environment. There were the host of freedoms associated with access to materials, activities, and areas, both within the school and within the community at large. Then there were the more personal freedoms such as the freedom of movement, of speech, of self-expression. There were freedoms associated with scheduling one's own activities, in one's own time, in a place of one's own choosing. There were also many freedoms which could be stated negatively: the freedom to leave a project momentarily without returning to find it destroyed or missing; the freedom to pursue a task without constant disruption; the freedom not to be continually abused or intimidated by fellow students as well as by adults. The list of potential freedoms seemed boundless.

Second, specific freedoms could be maintained only so long as they were accompanied by associated responsible behaviors. The Community School used many resources outside the school. However, when some of our children abused those resources we stood in danger of becoming less welcome guests. There have been many people who volunteered to conduct interesting projects at the school. However, when volunteers were verbally or otherwise abused they were less likely to volunteer their services in the future. When the destruction of property became rampant it was no longer possible to give children free access to equipment and materials. Thus, the abusive and destructive behavior of the children was eroding the freedoms within the Community School, and it was doing it in a way more certain than adult authority could ever do.

Another consideration was that all the children at the school were losing freedoms because of the irresponsible behavior of a minority of students. Even more frustrating was that the younger children were beginning to emulate the most aggressive, abusive, authoritarian behavior of their older peers. It seemed that the adults had relinquished their authority to the most aggressive children who were in turn shaping all the children to become progressively more insensitive toward each other.

The question that came sharply to focus from these discussions was how could a school develop and maintain the kind of responsible behaviors that would allow freedoms within the school to be maximized for all children. In the past we gave our children a school environment with a maximum of freedom from adult authority and assumed that their actions would rise to meet the occasion. They didn't. The opposite effect occurred and the result was a hostile, unproductive atmosphere. Therefore, we decided to try the opposite strategy. That is, we could award specific freedoms to individual children only as they began to demonstrate the appropriate responsible behavior.

Before attempting to put this new strategy to use on a school-wide basis we decided to try it first in a more limited way. The children had no previous systematic instruction in mathematics and their skills in this area were correspondingly limited. Therefore, we decided to examine the effect of making math class time available for students to use as free time dependent on their maintaining satisfactory rates in the acquisition of mathematics skills.

Setting

The Colorado Springs Community School has 40 students ranging in age from 4 to 14. It is located in a reasonably spacious building which has roughly ten rooms or areas of varying dimensions and a playground area outside. Children have free access to all areas except during classes. This study took place in the math area which contained five tables. Children were free to sit wherever they wanted to in the room.

Subjects

Eighteen children, 9 to 14 years old, were in the first class which met from 9:30 to 10:30 AM, Monday, Tuesday, Thursday, and Friday. Sixteen children, 4 to 8 years old, were in the second class which met from 10:30 to 11:30 AM on the same four days. Ten children in each class who were enrolled for the entire school year, and were working in the math program described below, served as subjects. Most of these children had no systematic instruction in mathematics for the previous two years and their skill level in this area was relatively low.

Materials

The math materials for this study were adapted from Patrick Suppes, *Sets and Numbers (Books 1-4, 1962)*. The material was transcribed onto ditto sheets, each sheet containing one or two concepts. Every new type of problem or new format was defined as a concept. Quizzes were written for each ten concepts directly corresponding to the problems on the worksheets. The students were required to get eight out of ten problems correct on each quiz before moving on in the program. Each child was allowed to work at his own pace and at his own skill level. There was no group instruction in math at any time during the school year. Throughout the entire study the children were required to have their work checked and correct any errors before leaving class. The time which the class began and the checkout time for each student was recorded each day.

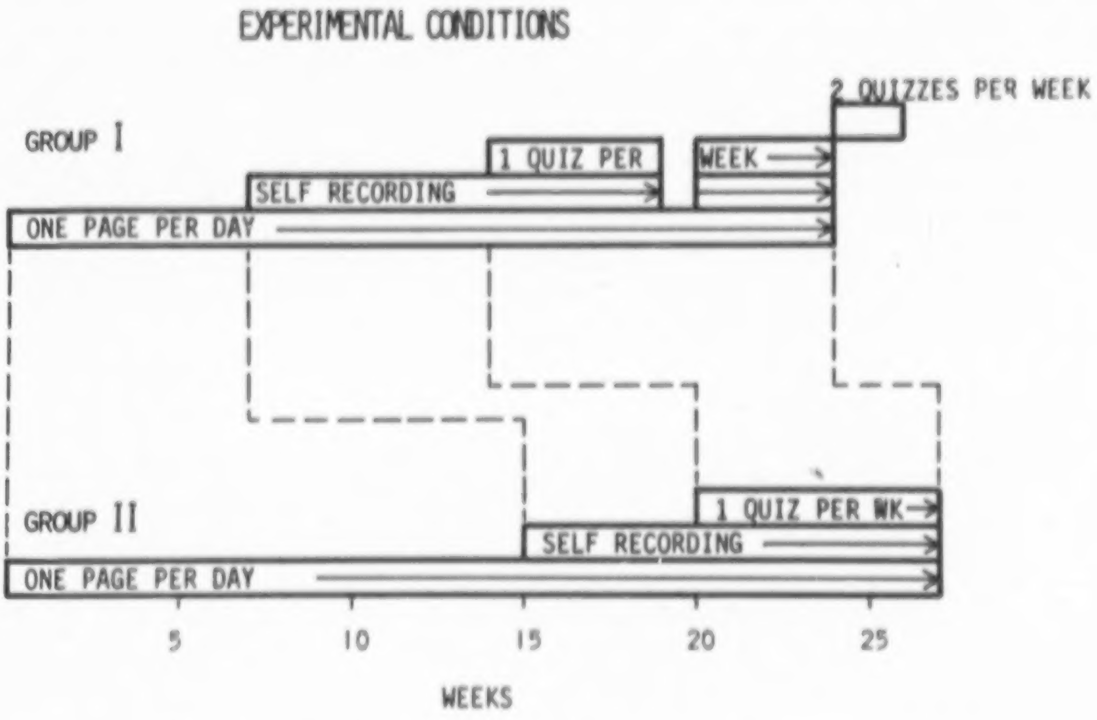

EXPERIMENTAL CONDITIONS
GROUP I
2 QUIZZES PER WEEK
1 QUIZ PER
WEEK
SELF RECORDING
ONE PAGE PER DAY
GROUP II
1 QUIZ PER WK
SELF RECORDING
ONE PAGE PER DAY
5
10
15
20
25
WEEKS

Fig. 1. The basic research design was a multiple baseline across groups.

Research Design

The basic design of this study was a multiple baseline across two different groups of subjects. To provide additional inferential power, a one-week reversal of procedures was included for the older children (Group I) in week 19. The sequence of experimental conditions is presented in Figure 1. Note that each new condition is added to, and includes, the previous conditions.

Experimental Conditions

One Page Per Day Minimum. During this condition students were required to come to class and complete a minimum of one page at which time they were free to leave.

Self-recording Procedure Plus One Page Minimum. Students were provided with charts in their math folders. Each time a child passed a quiz he colored in the appropriate block on his chart, as shown in Figure 2. During this condition students were still required to complete a page of work before leaving class.

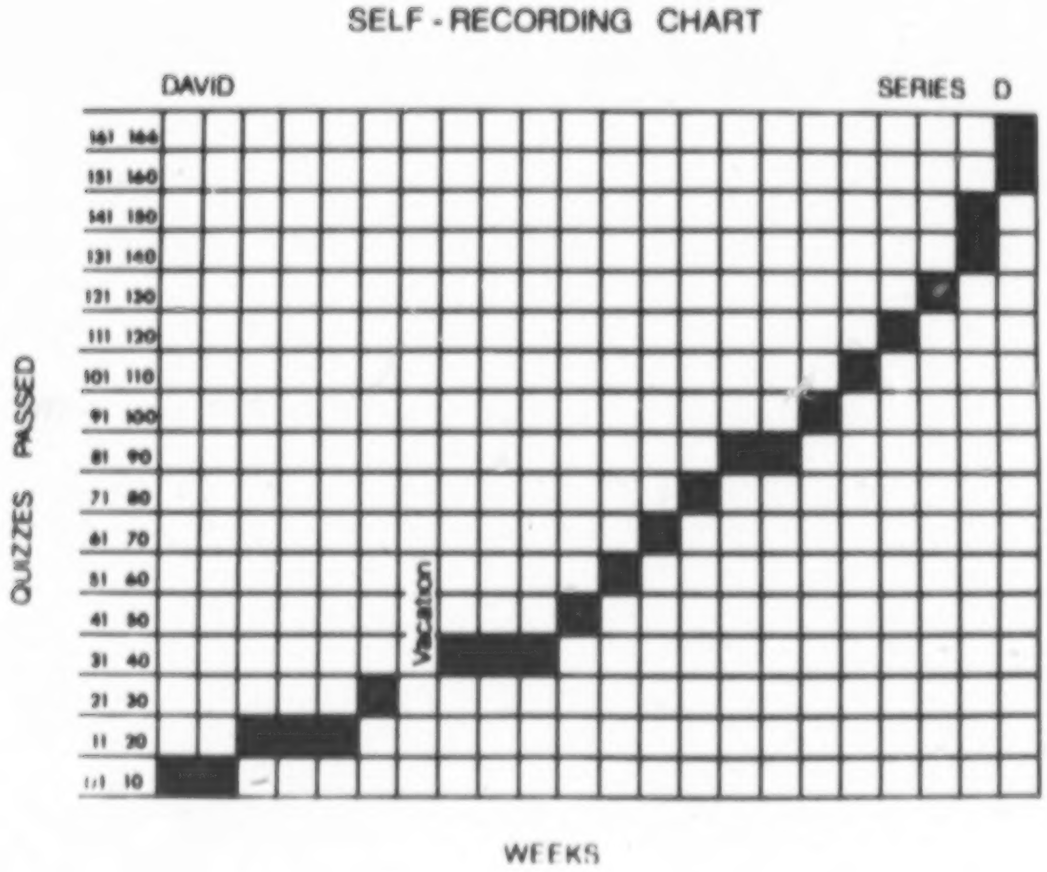


Fig. 2. Self-recording charts as shown here were used by the students to track their rate of progress through the math program.

One Quiz Per Week. During this condition the students were free to leave class as soon as they completed their one page a day minimum requirement and passed at least one quiz each week. If they failed to pass a quiz in any week they were required to stay in math class for the entire hour during the following week. This included students who were absent. They continued to chart their own progress through the math program.

Two Quizzes Per Week. Students in Group I were told that they did not have to attend class at all if they passed one quiz on Tuesday, and another on Friday. Students who elected this system were given a packet of materials containing ten concepts on Friday which was to be completed in time to take a quiz on Tuesday morning. If the Tuesday quiz was passed, they received the next packet of materials which was due on Friday morning.

They could, if they wished, come to class to receive instruction or ask questions. The only requirement in addition to passing the two quizzes a week was that they not disturb other students who were working in class. This condition was offered for two weeks to Group I only.

No Minimum Requirement. This condition was run for Group II students only during the first three weeks of the school year (prior to Session 1, Figure 1). In this condition students had to come to class but they could do as little or as much as they cared to and leave as soon as they desired.

Questionnaire

Because the time necessary to conduct actual preference experiments was lacking, a questionnaire was administered to both groups of students regarding their preferences for the various experimental conditions. The older children did the questionnaire independently. Since many of the younger children lacked the reading and writing skills necessary to fill it out by themselves the questionnaire was administered to them individually by an adult volunteer.

Reliability

The primary behaviors measured in this study, rate of passing quizzes and number of pages of work completed, both leave permanent records. The scoring of quizzes was checked by an independent observer who was not aware of the experimental conditions from which the quizzes came. One out of five quizzes was checked in this fashion for each phase of the experiment. The mean reliability was 97% agreement with a range of 90% to 100%. Percent agreement was calculated as agreements times 100 divided by agreements plus disagreements.

RESULTS

Figure 3 presents the total number of quizzes passed each week in both groups in all the experimental conditions. Each group had ten students. During the one page per day minimum condition both groups averaged about five quizzes passed each week. That is, it took a student approximately two weeks to complete a ten concept unit of work and pass the associated quiz.

During the next condition the students were introduced to the self-recording charts. Although they were interested in them and usually eager to mark in the appropriate block when they passed a quiz, their rate of progress in the math program remained basically unchanged.

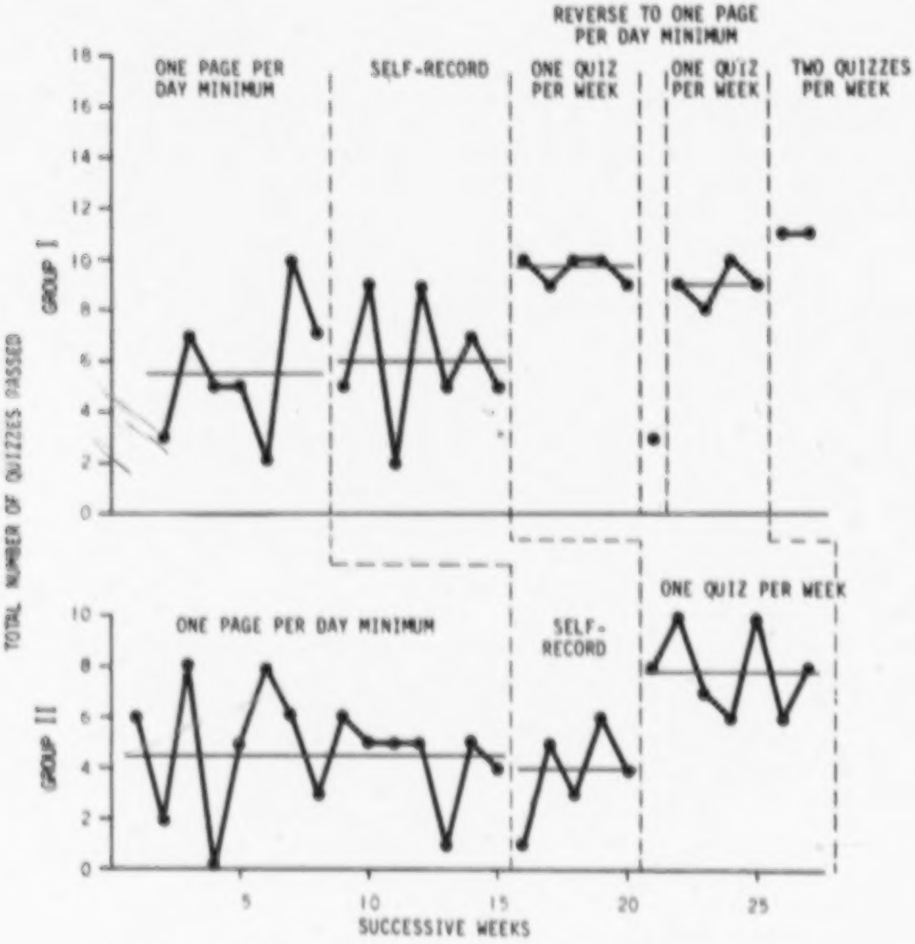


Fig. 3. The mean number of quizzes passed each week of school for Group I and Group II in all the above conditions.

There was an immediate and sharp increase in the rate of progress in both groups during the one quiz per week condition. The Group I students increased from a mean of 6.0 quizzes passed per week to a mean of 9.7. Group II students increased from 4.0 to almost 8.0 quizzes passed per week.

When the conditions were reversed for one week back to the one page per day procedure, Group I students returned to a rate even lower than their previous baseline rate. When the one quiz per week procedure was reinstated the rate of passing quizzes rose again to about its former rate during that condition.

When the two quizzes per week condition was begun for Group I, six students elected this procedure immediately. They received their packets

Friday, and five of the six students successfully completed their work and passed the Tuesday quiz. Only one of these five students, however, finished his Tuesday assignment on time and passed the quiz on Friday. The one successful student and one other student who had not previously elected this condition were eligible, and elected the two quiz per week procedure that Friday. Both students successfully completed the Tuesday and Friday assignments and passed the appropriate quizzes. On the second Friday, four students were eligible and elected the two quiz per week condition for the following week. Since this was the last academic week of the school year, the contingency which always applied to the subsequent week was no longer in effect. Consequently, the data from this final week were not included in the study.

During the first three weeks of the school year the younger children were allowed to do as little or as much math as they wanted to and could leave class as soon as they wished. It may be seen from Figure 4 that under the no requirement condition they did very little, an average of six problems each day. They were working at an average rate of less than one half problem per minute, and spending about 15 minutes in class each day. When in the next condition they were required to complete a minimum of one page each day before leaving class, they completed two and one half to three times as many problems. The upper axis in Figure 4 indicates that their rate of work also increased. Consequently, even though they were completing two and one half to three times the previous amount of work, they were only spending an additional five to seven minutes in class.

There were some correlated effects of the experimental procedures in addition to increasing the students' rates of progress in math. At the beginning of the school year it was necessary to round the children up for class each morning. During the one quiz per week condition students began showing up in class early to do their math. In fact, an average of 31% of the Group I students came to class early each day during the one quiz per week condition. No student ever came to class early in any of the previous experimental conditions. In the beginning of the school year no student ever voluntarily took any academic work to do at home. Figure 5 shows the total number of pages the Group I students voluntarily did outside the math class each week. During the self-recording condition two students occasionally took a page or two home and returned it completed. Three weeks into the one quiz per week condition, significant numbers of students began to take work home. This effect continued until the experimental conditions were reversed, when voluntary homework decreased to zero. When the one quiz per week condition was reinstated voluntary homework again increased.

The mean number of pages of work per day that each student completed in the last ten days of all experimental conditions is shown in Table 1. It may be noted that Student C in Group I did not respond to the experimental contingencies. All other students were influenced in varying degrees by the free time contingency.

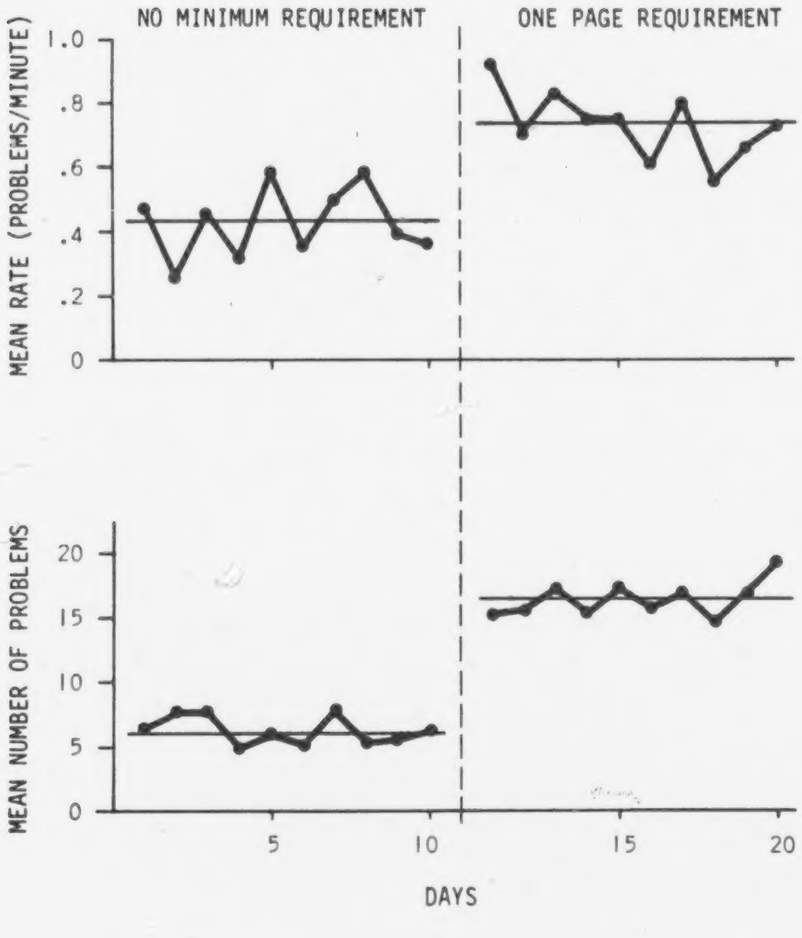


Fig. 4. The mean rate and number of problems completed by Group II students during the last ten days in each of the above conditions.

TOTAL NUMBER OF PAGES OF HOMEWORK COMPLETED (N = 10)

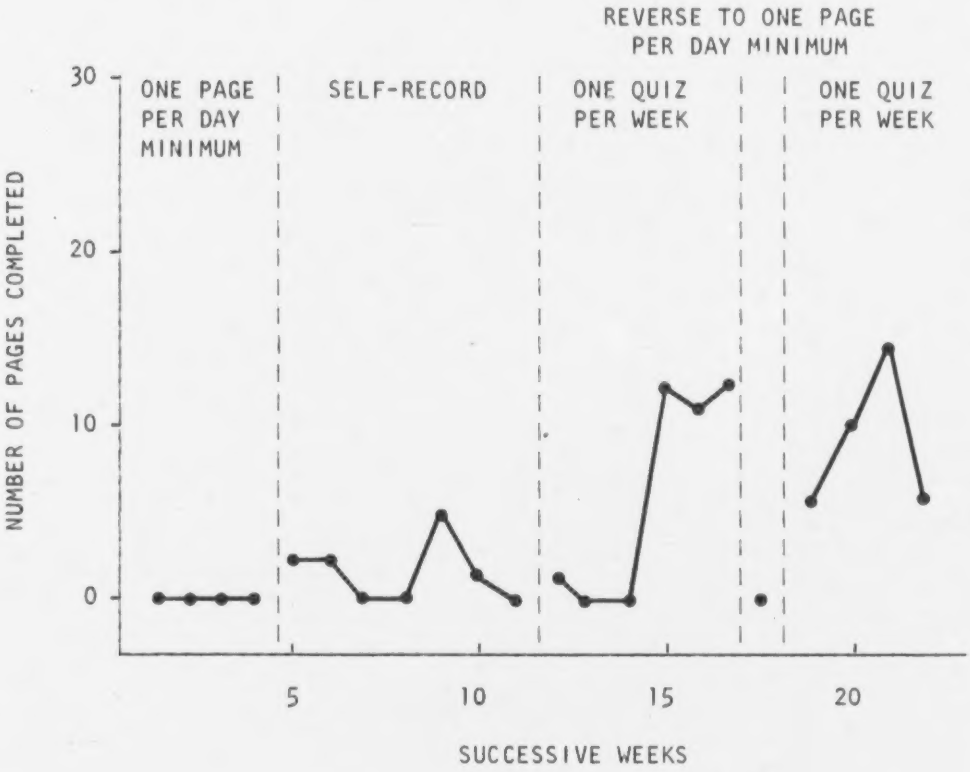


Fig. 5. The total number of pages of voluntary homework completed by Group I students each week in all experimental conditions.

TABLE 1

STUDENTS	1 PG/DAY	SELF-RECORD	1 QUIZ/WEEK	1 PG/DAY	1 QUIZ/WEEK	2 QUIZZES/WEEK
A	1.2	1.0	2.3	1.0	1.8	2.0 ✡
B	1.4	1.0	1.7	1.3	1.9	1.7
C	1.5	1.0	1.3	1.3	1.2	1.1
D	1.6	1.0	2.7	1.0	1.7	2.8 ✡✡
E	1.4	1.0	2.3	1.0	1.6	2.0 ✡
F	1.7	1.0	3.1	1.0	1.7	4.0 ✡✡✡
G	1.9	1.3	2.4	1.0	1.8	1.9 ✡
H	1.3	1.2	1.9	1.0	2.0	1.5
I	1.5	1.0	1.8	1.0	1.4	2.0 ✡
J	1.2	1.0	1.4	1.3	1.8	2.1 ✡

GROUP I

STUDENTS	1 PG/DAY	SELF-RECORD	1 QUIZ/WEEK
A	1.0	1.1	1.5
B	1.1	1.3	1.6
C	1.1	1.2	1.6
D	1.1	1.2	1.6
E	1.1	1.2	1.6
F	1.1	1.0	1.8
G	1.2	1.0	1.8
H	1.0	1.1	2.4
I	1.0	1.1	1.4
J	1.0	1.0	2.1

GROUP II

The mean number of pages per day of math completed by each student during the last 10 days in each experimental condition. The asterisks indicate the number of times each student in Group 1 successfully completed the quiz per one half week criteria needed to stay eligible for the two quizzes per week condition.

The results of the questionnaire are presented in Figure 6. When the older children chose between the two quiz per week condition and the one quiz per week condition, seven of the nine preferred to do more work and not go to class at all. These were the same seven children who had actually attempted (though not always successfully) that procedure. Of the two older children who generally preferred the less work, less free time condition, one was the student for whom the free time contingency had no effect. The younger students, on the upper axis, had never had the opportunity to elect the two quiz per week condition and their stated preferences regarding that differ from those of Group I. The last two preference comparisons indicate that there is a strong stated preference to meet relatively high work criteria when the alternative is to be in class the entire hour even without any minimum work requirements.

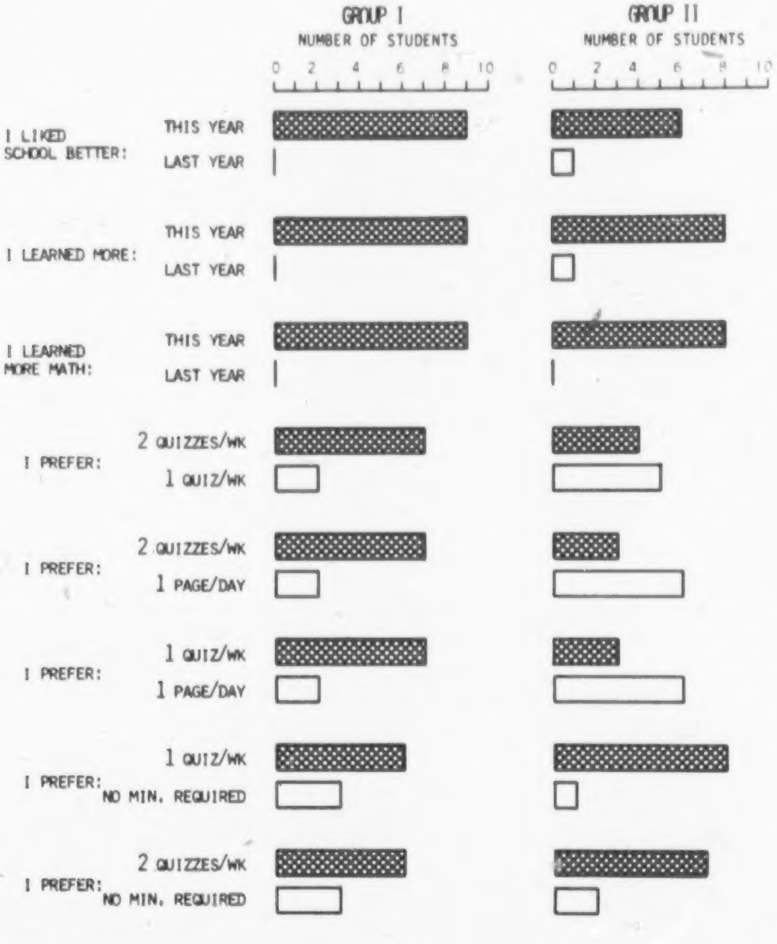


Fig. 6. The results from the questionnaire given to Group I and Group II students at the end of the school year.

DISCUSSION

The effects of the experimental procedures in this study were certainly desirable from a teacher's point of view. During the one page minimum requirement the students' rates of progress would allow them to complete about one normative year's work in one year. Their rates of work during the one quiz per week condition would allow them to cover two normal years of math training in one year. Students successfully meeting the two quizzes per week contingency were working at a rate three and one half to four times that of their normative peers. Correlated increases in self-help behaviors and increases in voluntary homework and promptness in coming to class produced a more pleasant, cooperative instructional climate.

The experimental procedures produced increases in the rate of progress for 19 out of 20 students. Most students spent their earned free time socializing with each other. The one student for whom the procedures were not effective had a long history of aversive social interactions with the other children at the school. This may be related to the ineffectiveness of the free time reinforcer for this child. One could allow children to specify alternate reinforcers. One could also attempt to operate directly on this child's social interaction.

In addition to examining the effect of the experimental procedures on the rate of work, it also seemed desirable to develop classroom procedures which would take into account the preferences of the children. The results of the questionnaire presented in Figure 6 indicated for most children a stated preference for high work loads if these are matched by increasing opportunities for free time.

These data refer primarily to academic behavior. However, creating a situation in which it was in the direct interest of each student to progress through the math program also produced a wide variety of self-help behaviors; from procuring one's own pencils and materials to taking some work to do at home. Unfortunately we lacked facilities to objectively measure all these additional effects. The teachers would sum it up by saying that the students were beginning to take initiative for their own academic progress.

The strategy of awarding freedoms contingent upon associated responsible behaviors is applicable to social as well as academic areas. We are in the process of developing some mechanisms to try it out on a school-wide basis. It seems that this strategy has three advantages. First, it insures maximum prerogatives to those children whose behavior merits them. Second, it protects the freedoms of those who have earned them from being endangered by the behavior of those children who have been characteristically abusive. Most important, this strategy has the potential of moving all the children toward being the kind of thoughtful, productive, independent individuals whom we very much want to encourage.

Those who are interested in conducting free schools will come, in time, to recognize, as we did, that there is more to the creation of a free environment than removing adult supervision. Freedom can exist in a school *only* to the extent that the students in that school are responsible. The potential freedoms available within a school can produce and maintain that responsible behavior but *only* when access to those freedoms is made contingent upon that responsible behavior.

REFERENCES

Suppes, Patrick. *Sets and numbers*, Books K through 6. New York: Random House, Inc., 1962.

* * *

ACHIEVEMENT TESTS VS. DIRECT & DAILY MEASUREMENT[1]

marie d. eaton
tom c. lovitt

In September or May many children in elementary schools throughout the country are given achievement tests. The scores from these tests are used for three major purposes: *placement*, *communication* and *evaluation*. In this paper the use of achievement test scores for these purposes will be challenged. Direct and daily measurement will be offered as an alternative method of assessing performance.

Placement

Achievement test scores are often used to place children in one class or another (e.g., second or third); in one reading grade level or another (e.g., 2-2 or 3-1). If, for example, a pupil's subtest scores at the beginning of the year on an achievement test are second grade-six months (2.6), he could be placed in a reader that supposedly corresponds to the second half of the second year (2-2).

Such a placement could, for a number of reasons, be highly inappropriate. One factor that might account for the mismatch between the pupil and program could be that the achievement test was given only once. It could be that the child was a very fortunate guesser on the day of the test and received a score much higher than his true ability. Conversely, the pupil may have been ill or upset on the day of testing and obtained a score far below his real competence.

A second factor that could account for an incongruous union between pupil and program is that the achievement test sampled one set of behaviors and the teacher and her reading program stressed another set of behaviors. The achievement test could have assayed such reading and language skills as reading words in isolation, spelling, word definitions. Meanwhile, the important reading elements, as far as the child's teacher and her favorite reading program were concerned, were reading orally in context and answering comprehension questions.

---

[1] This research was partially supported by a grant from the United States Office of Education, Bureau of Education for the Handicapped, Grant No. OEG-0-70-3-916(607).

Following are the procedures used to place a child in a reader using direct and daily measurement. In this example, the decision was made to focus instruction initially on oral reading. Later, when the pupil could orally read at a competent rate in an appropriate level (e.g., 100 correct wpm in a 3-1 reader), instruction would concentrate more on silent reading and comprehension.

Since oral reading was measured, the behavior of concern, the assessment was direct. For several days the lad's teacher required him to read from a number of books from three reading series--Bank Street, Lippincott and Ginn. Thus, his performance was measured daily.

Figure 1 presents the correct and error rate ranges throughout the assessment period for each book. The hatch mark through each range points out his median performance. The asterisk indicates which grade level from each reading series was selected for the child's reading program.

When direct and daily measurement procedures are used to place children for instructional purposes in particular reading programs, the probability is great that a congenial merger between pupil and program will be realized. One reason for this confidence is that when several assessments of performance are obtained before a placement decision is made, unusual ratings will be tempered by more normal scores. A second reason to support such a presumption is that when the behavior of concern is measured directly, when the diagnosed behavior becomes the instructed behavior, a mismatch between pupil and program is impossible.

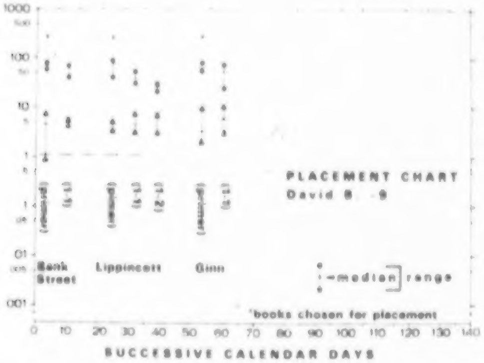


Fig. 1. Placement data for one child. Correct and error rate ranges and medians from numerous reading levels in three reading series were plotted. The asterisks indicate which books were chosen for initial instruction.

Communication

Achievement test scores are often used to report a child's progress to his parents and his teacher for the coming year. Because of the criticisms mentioned previously regarding achievement tests, such a method for reporting can be quite misleading. A measure that is taken only once or twice a year that doesn't necessarily measure what the child is being taught, offers a flimsy basis for communication.

Following the administration of an achievement test at the close of a school year, a pupil's scores in the reading-language arts subtests might indicate that his word recognition score was 2.6, his spelling score 2.8, and his word meaning score 2.4. Whereupon these statistics could be passed on to the pupil's parents and his teacher for the coming year. Such information, furnished only at the end of the year, would not reveal to the parents how much progress their child had made throughout the year. Neither would the parents have information about several other skills that were being taught their child, but were not evaluated by the test. Next year's teacher, if she consulted such a dossier, would be equally baffled. She would not know exactly which skills the pupil had been working on, which ones he had mastered, or which instructional techniques were effective.

To a great extent, this communication bewilderment can be overcome if pupil performance is measured directly and daily. Again, in the area of reading, if the pupil's reading program comprised oral reading, answering comprehension questions and writing resumes about the stories, measures on those behaviors would be sent to the parents and receiving teacher. Furthermore, such communication would describe the pupil's attainments in these skills over an extended period of time. Parents receiving such information would be enlightened as to their child's accomplishments. The receiving teacher, so notified, would know exactly which skills were being attended to, which materials and procedures were being used, and, perhaps, which instructional techniques had been effective.

Direct and daily measurement facilitates teacher-pupil communication. If the child is involved in looking at or charting his own performance scores each day, he becomes aware of his progress; often being able to see what is "going on" is motivating for a child. This alone can inspire children to improve their performance. Other children, if they "see" their achievement worsening from day to day, request the teacher to help them.

Evaluation

Achievement test scores are also used for purposes of evaluation. These scores are used to compare various methods, such as the phonics or the sight word method; materials, such as the Bank Street or Ginn readers; tactics, such as flash card drill or tachistoscopic practice.

The same criticisms of achievement tests (indirect and infrequent), when used for placement and communication, are true when such tests are used for evaluative purposes. For instance, if a teacher uses the achievement test methodology to compare the effects of Ginn and Lippincott readers, she could first establish two groups of children, a Ginn and a Lippincott group. Then she could pretest the groups with an achievement test, then the treatment periods could be scheduled, followed by a posttest on the same or an equivalent achievement test. She might then determine the mean and standard deviation scores of the groups, run a t-test and announce the winner.

In such a scheme, if one reading series seemed to produce better results than the other, it could be due more to the fact that the behaviors taught by that method were more congruent with the items sampled by the achievement test than was the case with the other method, than due to the excellence of that particular system. Also, the fact that the achievement test was given only twice--at the beginning and end of the treatment--can provide misleading information. It could be that the results at the end indicated that gains occurred for both groups. For one group, however, this gain may have been gradual, throughout the treatment; for the other, it could have happened during the first days of the treatment.

When direct and daily measures of performance are used to make similar comparisons, the results are more valid and certainly more individually relevant. If the teacher desired to compare oral reading rates, for example, from the Ginn and Lippincott readers, she could simply have the pupil read from both texts for several days and chart correct and error rates for each session. Throughout this period of assessment she could, by observing the data, determine which reader was most suitable for the child. She could make daily, weekly, or monthly across-reader comparisons as to correct and error rates. Such an evaluative scheme is assuredly preferred to the achievement test strategy since the data pertain to individuals not groups; are obtained daily, not before and after treatment; and bear a direct not an indirect correspondence to the program being used.

ACHIEVEMENT TEST AND DIRECT-DAILY MEASUREMENT: A COMPARISON

The data which follow are achievement test scores for two boys and their direct and daily measurement throughout the school year, 1970-1971. Paul L. and John were referred to the Curriculum Research Class of the Experimental Education Unit because of reading problems. They were in a class with four other pupils. In September of 1970, all the students in the class were given the WIDE RANGE ACHIEVEMENT TEST *(WRAT)* (Jastak, Bijou and Jastak, 1965) and the METROPOLITAN ACHIEVEMENT TEST *(MAT)* (1959, 1961, 1965). Both tests were given again in the spring of 1971.

Table 1 shows the spring and fall scores for the class in the two achievement tests. Also indicated in the table is the form of the achievement test that was given, the names of the books the pupils were placed in, and the grade levels of these books in the fall and spring.

TABLE 1

READING -- GRADE LEVEL SCORES

1970-1971

CHILD	METROPOLITAN ACHIEVEMENT TEST		WIDE RANGE ACHIEVEMENT TEST		ACTUAL PLACEMENT		
	FALL	SPRING	FALL	SPRING	BOOK	FALL	SPRING
PAUL L.	1.0 $^{P^2}$●	2.7 ^{E}	2.7	2.5	LIPPINCOTT PALO ALTO BANK STREET	--- BOOK 3 PRIMER	2-2 --- 2-2
PAUL R.	2.6 ^{E}	4.0 ^{I}	2.5	2.5	BANK STREET SCOTT FORESMAN	2-1 2-1	3-1 4-1
JOHN	1.1 ^{P}	4.0 ^{E}	KG 7	8.7	LIPPINCOTT PALO ALTO	PRE-PRIMER ---	2-1 BOOK 11
FRED	1.6 ^{E}	2.5 ^{E}	1.9	1.9	MERRILL BANK STREET	--- PRIMER	BOOK 5 2-2
JENNI	2.4 ^{E}	6.8 ^{I}	4.4	4.8	LIPPINCOTT LAIDLAW MCMILLAN	3-1 3-1 3-1	BOOK 5 BOOK 5 BOOK 5
PHIL	NONE LOW ENOUGH	1.7 ^{P}	KG 5	1.9	LIPPINCOTT BANK STREET	PRE-PRIMER ---	1-1 1-1

● = INDICATES WHICH BATTERY WAS GIVEN

P = PRIMARY BATTERY

P^2 = $PRIMARY^2$ BATTERY

E = ELEMENTARY BATTERY

I = INTERMEDIATE BATTERY

Throughout the year direct and daily data were kept on Paul's and John's progress in oral reading. These data, when compared with the achievement test scores, often suggest contrary interpretations of accomplishment. The bases for some of these conflicting notions will be discussed later.

All the data for the direct and daily measurement were taken in a small room off the main classroom. In the room were a table, two chairs, a tape recorder and some books. Only the teacher and the child were present during the reading sessions. Each child read for five minutes from two readers each day. The teacher counted correct and error words orally read and the reading was timed with a stopwatch. Errors consisted of substitutions, additions, and omissions. All errors were corrected by the teacher after a 5-second pause to allow for self-correction. If the child corrected himself before the teacher supplied the words it was not counted as an error.

Paul

Paul L. was an eleven-year-old boy. When he was given the MAT (Primary 2 Battery) and the WRAT in September, his reading scores were 1.0 and 2.7 (see Table 1). In September when the teacher used direct and daily measurement for diagnosis, he was placed in the Palo Alto Book 3 (first grade level) and the Bank Street primer. Thus, the MAT score was more commensurate with his actual placement than the WRAT. In September, Paul's median correct rate was 29.6 words per minute (wpm) in the Bank Street primer with a median error rate of 2.0 wpm (see Figure 2).

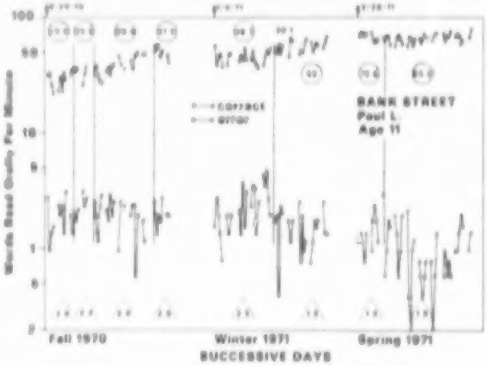


Fig. 2. Direct and daily measurement of words read orally in the academic year 1970-1971 for Paul L. Any breaks in the data indicate absences, weekends or vacations. The vertical lines indicate when different instructional procedures were put in effect. The circled numerals above the correct rate plots indicate median correct rates for each phase and the numerals within the triangles indicate median error rates. Two medians were figured for conditions which were separated by vacation periods. The arrow and date at the top of the chart indicate the days data were first taken during each quarter.

In May, when the achievement tests were given again, Paul scored 2.7 on the MAT reading subtest and 2.5 on the WRAT. From fall to spring his MAT score rose 1.7 of a grade level and his WRAT score fell .2 of a grade level. In May, Paul read in the 2-2 reader of the Bank Street and Lippincott series. At that point, both achievement test scores agreed rather closely with the direct and daily measurement. Neither achievement test score, however, revealed, as did the direct and daily measurement, that Paul's correct rate in May had risen to a median of 65.8 wpm and his error rate had dropped to 1.0 wpm. Thus, throughout the year, Paul made progress not only through the materials--from primer to 2-2 in Bank Street--but also showed a definite improvement in correct and error rate. From September to May he was reading more difficult material faster and more accurately.

Paul's direct and daily measurement also clearly show how this method of assessment enabled his teacher to evaluate various instructional techniques. As indicated in Figure 2, Paul's improvement was not steady throughout the year. At a number of times between September and May, Paul's performance began to deteriorate; his correct rate fell and his error rate rose. Because the teacher was obtaining direct and daily measurement, she saw these patterns as they developed and was able to make instructional or contingency alterations immediately. She did not have to wait until May to discover that the pupil was not progressing satisfactorily.

In the winter quarter, such a performance decline was noted. The vertical line in the figure from winter, 1971, indicates where a decision to change procedures was made. At this point a 30:1 point contingency was instituted. For every thirty correctly read words, Paul received one point, redeemable for one minute of free time. Following this procedural change, Paul's performance improved; his correct rate rose while his error rate declined.

Because the teacher was taking direct and daily measurement on oral reading, she was certain that those data meant progress in that skill. The data did not reveal, necessarily, any information about his ability to perform on an achievement test. This point is especially relevant when considering the scores from the WRAT. Paul's reading scores from September to May in the WRAT showed a drop of .2 of a grade level, yet his daily data showed that he made considerable gains.

The reading subtest in the WRAT requires the child to read a list of words. The direct and daily measurement pertained to Paul's reading words in context. Reading word lists is certainly a different task than reading a story, so a score based on reading word lists might have little relationship to a score on reading a text.

John

John was a nine-year-old boy. In September, when the achievement tests were first given, his score was 1.1 on the MAT (Primary Battery) and at the seventh month of kindergarten on the WRAT (see Table 1). At that time John was placed in a Lippincott pre-primer on the basis of direct and daily measurement taken in a number of readers.

Figure 3 shows John's direct and daily data from fall and spring quarters. During the first three weeks of school John's median correct rate in the Lippincott book was 45 wpm with an error rate of 1.8 wpm. His reading fluctuated widely from day to day; correct rate scores ranged from 21 to 100 wpm and error rate scores from .3 to 9.0 wpm. In May, John was reading in the 2-1 Lippincott reader; his median correct rate was 75 wpm and his median error rate was 2.2 wpm. As the chart shows, his performance, insofar as correct and error rate was concerned, was much more stable in the spring than in the fall.

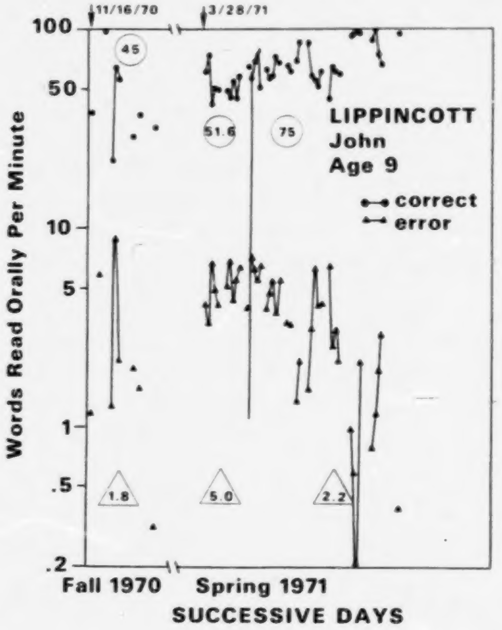


Fig. 3. Direct and daily measurement of words read orally in the fall and spring quarters (1970-1971) for John. The vertical line in the spring quarter indicates that a procedural change was made.

John's data show even more clearly than Paul's how discrepant a child's achievement test scores can be from his direct and daily measurement. John's progress, as indicated in Figure 3, was impressive. From September to May he was reading more advanced material, his correct rates were more stable, his correct rates higher, and error rates lower. In spite of this improvement, however, he did not progress eight years in reading ability in one year's time as the WRAT scores would suggest (see Table 1). As previously mentioned, the reading test on the WRAT asseses the skill of reading isolated words, while the direct and daily measurement pertained to reading words in context. If John's teacher for the following year were to place him on the basis of his WRAT score, she might be tempted to begin him in a junior high reader, when he, in fact, should read from a second grade text.

DISCUSSION

Direct and daily measurement not only provides placement, communication and evaluative information, but also furnishes some valuable clues for remediation. When the direct and daily measurement method is used for initial placement purposes, the teacher may be supplied clues regarding which tactic to use in order to assist his pupil's reading. Using the direct and daily method, a teacher can easily observe which types of reading errors are being made. It may be revealed that the child reads in a jerky manner, that he can't decode words that begin with *cl* or *br*, that he substitutes one noun for another, or that he confuses pronouns. When specific remediation pinpoints can be identified early, remediation and subsequent improvement can be quickly realized.

By contrast, achievement test scores provide little if any assistance to the teacher when it comes to selecting remediation strategies. The same criticisms expressed earlier of achievement tests, that they provide indirect information and that, only infrequently, preclude their being used for diagnostic purposes.

In reading, when an achievement test is used to indicate progress from one period of time to another, only a grade level change is provided (e.g., from 2.6 to 4.0). As discussed, this grade level advancement may not correspond to the reading behaviors that were being taught. When direct and daily measurement is used to evaluate progress, grade level changes pertain to the program which the child is using. In addition to providing grade level information, direct and daily measurement shows a pupil's correct and error rate scores over a period of time. The ratio between these two rates at any given time is an indicator of the quality of his performance. Thus, speed of reading correctly and incorrectly as well as accuracy are shown.

REFERENCES

Jastak, J. R., Bijou, S. W. and Jastak, S. R. *Wide range achievement test.* (Rev. Ed.) Wilmington, Delaware: Guidance Associates, 1965.

Metropolitan achievement test. Primary Battery. New York: Harcourt, Brace and World, 1959.

Metropolitan achievement test. Primary 2 Battery. New York: Harcourt, Brace and World, 1961.

Metropolitan achievement test. Elementary Battery. New York: Harcourt, Brace and World, 1965.

* * *

ACQUISITION OF PHONETIC SOUNDS BY PRESCHOOL CHILDREN

I. EFFECTS OF RESPONSE AND REINFORCEMENT FREQUENCY

II. EFFECTS OF TACTILE DIFFERENCES IN DISCRIMINATIVE STIMULI[1]

virginia i. massad
barbara c. etzel

The initial study in Experiment I compared preschool children's acquisition of pointing to a letter in response to its phonetic sound during conditions of a high ratio of required responses and delivered reinforcers to instructions or cues (frequent condition) with a low ratio of the same (infrequent condition). The first control study compared the effects of a high ratio of delivered reinforcers to instructions with a low ratio of the same while required responses as well as number of instructions were held constant. The second compared the effects of a high ratio of required responses to instruction with a low ratio of the same while delivered reinforcers were held constant.

In the three studies, each subject served as his own control (i.e., a discrimination between two letters was acquired during conditions of high ratios and another discrimination was acquired between two different letters during conditions of low ratios. Acquisition with these procedures was replicated for each subject with a second set of letter pairs. The conditions during which the pairs of letters were trained and the sequence of training either the high or low ratio discrimination first during a session was controlled between subjects.

---

[1] This study was supported in part by an EPDA Grant, Training Teachers of Teachers (TTT), in Behavior Modification in Early Childhood Education OEG-0-70-1820(721) granted to the Department of Human Development, University of Kansas. Additional support was given to the experimenters through the University of Kansas Center for Research in Early Childhood Education, in association with the United States Office of Education (CEMREL, Inc., Grant Numbers OEC 3-7-070706-3118 and OEC 0-70-4152 607).

> The results of Experiment I showed that acquisition occurred more rapidly and with fewer errors under the frequent condition than the infrequent condition. The results of both control studies supported the conclusion that the high ratio of responding was a more important variable in Experiment I than the high ratio of reinforcement as used in this study.
>
> Experiment II compared the acquisition of two letter-sounds under conditions of a tactile response to a sandpaper letter and a tactile response to a painted letter. Both conditions incorporated a high ratio of required responses and reinforcers to instructions. Each subject was his own control (i.e., acquisition of one letter-sound was with the condition of a sandpaper stimulus and the other with a painted stimulus). These acquisitions were replicated with each subject with two additional letters. The conditions during which the letters were taught and the sequence of teaching either sandpaper or paint first during a session was controlled between subjects. The results showed that acquisition was more rapid, with fewer errors with tactile responses to sandpaper than to paint letters.

Educational and psychological literature has not yielded a large number of studies to guide teachers in their choice of procedures or materials when teaching conceptual skills. Intuition rather than data generally guides educators in selecting a particular procedure or a piece of equipment. Those few studies that do have a more direct application to the *methods* area are those that have shown the effects of a *mediating* verbal response on measures of: rate of learning (Weir and Stevenson, 1959); memory (Hagan and Kingsley, 1968); and accuracy of responding (Lovitt and Curtiss, 1968). Two studies, also concerned with procedures and materials compared active participation to passive listening of kindergarten children (Blank and Frank, 1971); and rough with smooth stimuli (Gliner, 1967).

Many preschools, kindergartens and special education classes have used Montessori materials for years, but comparisons that have been made on the construction of the materials and the method of using them are so gross that statements cannot be made concerning what piece of equipment, used in what way, will or will not result in a child's acquisition of an academic skill. It still remains necessary for learning specialists to specify those variables.

The present research was carried out to serve as one type of model for studying specific variables in methods and materials research and as an example of a design that can more efficiently handle individual differences among subjects' learning rates. It involved preschool children's acquisition of pointing to letters in response to their phonetic sound production by the experimenter. This acquisition was compared during different teaching procedures and with different teaching materials.

Experiment I compared the frequency (frequent versus infrequent) of a child responding (and reinforcement) while the amount of information given under both conditions was held constant. Two control studies were then carried out. In the first control study the number of reinforcers given to the child were varied (frequent versus infrequent reinforcers), but information, cuing the responses and number of responses were all held constant. The second control study varied the number of teacher cues and subsequent child responses (frequent cue-responses versus infrequent cue-responses) while holding information and reinforcement constant.

Experiment II compared the acquisition of pointing to letters in response to their phonetic sounds using tactile responses to sandpaper with acquisition using tactile responses to painted letters.

Table 1 summarizes the variables studied across the two experiments and the two control studies. The upper box separates the two main variables studied in Experiment I (experimenter-cue and subsequent subject-response, and reinforcement) into frequent and infrequent conditions for each variable. As indicated, Control Study A was designed to investigate frequency of reinforcement and Control Study B to investigate frequency of experimenter-cue and subsequent subject-response.

The lower table summarizes the four studies by the type of variable being investigated and the number of times it was presented. Brief examples of each of the variables manipulated in the training sessions are noted under *examples and conditions*. In all four studies, each subject received both independent variables.

METHOD

Subjects and Setting

A total of 25 children from the Edna A. Hill Child Development Preschool Laboratories at the University of Kansas and the Lawrence Presbyterian Church Preschool participated in these studies.[2] They ranged in age from three years, one month to four years, eight months. All children had in-

---

[2] The authors wish to thank Mrs. Josephine Anderson for her kind assistance in carrying out this project with their children, at their school.

TABLE 1

DESIGN OF EXPERIMENT I AND CONTROLS A AND B☆

EXPERIMENTER-CUE AND SUBJECT-RESPONSE

		Frequent	Infrequent
REINFORCEMENT	Frequent	*Experiment I* *Control A* *Control B*	*Control B* $N = 2$
	Infreq.	*Control A* $N = 4$	*Experiment I* $N = 11$

☆*Each study (example:* CONTROL A) *coatains the same subjects in both cells.*

FREQUENCY OF EXPERIMENTER'S PROCEDURES & SUBJECT'S RESPONSES PER TRAINING SESSION ACROSS GROUPS & STUDIES

Study	Independent Variables✱	Frequency of Experimenter's Informative Statement Per Letter Per Session✱✱	Number of Paired Experimenter-Cues & Subject-Responses Per Session✱✱✱	Frequency of Reinforcers Per Session ✱✱✱✱
Exp. I	Fr-Resp.	5	5	5
	In-Resp.	5	2	2
Control **A**	Fr-Reinf.	5	5	5
	In-Reinf.	5	5	2
Control **B**	Fr-Resp.	5	5	5
	In-Resp.	5	2	5
Exp. II	Sandpaper	5	5	5
	Paint	5	5	5

✱*Variables:* Fr-Resp. = Frequent Response
Fr-Reinf. = Frequent Reinforcement
In-Reinf. = Infrequent Reinforcement
In-Resp. = Infrequent Response

Examples of Conditions:

✱✱Informative Statement: Experimenter, "The sound of this letter is "bů." (bů as in book)

✱✱✱Cue: Experimenter, "Point to bů and say bů."
Response: Subject points and says, "bů."

✱✱✱✱Experimenter gives praise and token.

dividual training sessions and test probes each day they attended school. These were held in experimental rooms in close proximity to the preschool classes.

Materials

The materials used were the Montessori lower case letters made of sandpaper (and paint for Experiment II). These letters were mounted on 4" x 6" red fiberboard. The pairs of letters trained were *m-c*, *b-n*, *a-s*, *t-r* and *h-c* (the latter being a substitute pair). The individual letters were selected on the basis of the frequency of their occurrence in early reading words and the pairings were made on the basis of differences (subjective) on a visual and auditory basis.

Pretest and Posttest

A pretest was administered to screen for prior knowledge of each pair of letter-sounds to be trained. The response required was one of pointing to the letters in response to their phonetic sound production by the experimenter. Each letter-sound was tested three times and any child who scored over 33% correct on any pair of letter-sounds was not included in the study. Each response on the pre- and posttests that the child made (regardless of whether it was correct or incorrect) was followed by a comment such as, "Good try," or "That's fine."

After acquisition of the letter-sounds, subjects were posttested with the same procedures and stimuli as those used in the pretest. The criterion for moving from the training sessions to the posttest was 100% correct responding for two successive days on the test probes given during the training sessions.

Reinforcement Procedures and Test Probes

During acquisition training of the letter-sounds all the subjects' correct responses were reinforced with a token (with backup toy available) plus a remark of praise or approval from the experimenter. Correct responses emitted during the daily test probes which immediately followed the daily training sessions were reinforced in like manner. The experimenter made no response to incorrect choices during the test probes, but simply went on to the next trial.

Reliability

Reliability checks were taken intermittently during the four studies by having an observer independently record the experimenter's training procedures and the subjects' responses during training, test probes and posttests. The average agreement was 99%.

EXPERIMENT 1

Training Sessions and Test Probes

After the pretest, eleven children were taught two pairs of letter-sounds, one pair under frequent conditions, the other pair under infrequent conditions. As Table I shows, both letters of the pair assigned to the frequent condition of training were presented individually with five experimenter-cues for responses to point (to the letter) and say (the sound of the letter); five subject-responses and five reinforcements. Each letter of the pair assigned to the infrequent condition was presented with two each of the same procedures. All other teaching information given to the subjects was the same under both frequent and infrequent conditions. The first experimental condition presented on any one day alternated between the frequent and infrequent conditions.

During the same session, and immediately following the training of each pair of letter-sounds, the child was given a six-item test probe on the two letters. Each test probe contained six trials (three presented in a random order for each letter-sound). Each test probe trial always involved the two letters that were being taught and an additional distractor (the letter *x* or *z*). These test probes had two forms (alternated between sessions) and differed in arrangement of the three stimulus choices.

Although the ratio of responses and reinforcement was five for each letter during frequent to two during infrequent conditions, the amount of experimenter information provided for the child was the same. There were ten identical informative statements during both conditions. The only difference between the two conditions was the number of experimenter requests for a subject to respond, the number of responses and the number of reinforcers.

All subjects continued this training until they met a criterion of two consecutive days of 100% correct responding on the daily test probes under both conditions, in addition to 100% on the posttest.

Design

In Experiment I an individual analysis with group controls was incorporated. The design is presented in Table II where the conditions are noted for each subject who served as his own control for the main independent variable under study (frequent versus infrequent). In the same session each child was taught two letter-sounds with frequent responses and reinforcement and two letter-sounds with infrequent responses and reinforcement. Therefore, this was an individual analysis design, in that each child experienced both frequent and infrequent conditions, with a replication (noted as Problem 2 in Table II) of these conditions with four additional letter-sounds. The subjects were grouped and the groups

TABLE II

DESIGN OF EXPERIMENT I FOR INDIVIDUAL SUBJECTS FOR PROBLEMS 1 AND 2

PROBLEM I					
Frequent Condition				Infrequent Condition	
Group	Subj. No.	Pairs of letters trained	Session when the frequent condition was trained first	Pairs of letters trained	Session number when the infrequent condition was trained first
I	1	r-t	odd	b-n	even
	2	t-r	odd	n-b	even
	3	r-t	even	b-n	odd
	4	t-r	even	n-b	odd
II	5	r-t	even	b-n	odd
	6	t-r	even	n-b	odd
	7	b-n	even	r-t	odd
	8	n-b	even	t-r	odd
	9	b-n	odd	r-t	even
III	10	a-s	even	b-n	odd
	11	b-n	odd	a-s	even

◀ PROBLEM 2 ▶			
Frequent Condition		Infrequent Condition	
Pairs of letters trained	Session when the frequent condition was trained first	Pairs of letters trained	Session number when the infrequent condition was trained first
s-a	even	m-c	odd
a-s	even	c-m	odd
s-a	odd	m-c	even
a-s	odd	c-m	even
c-h	odd	a-s	even
h-c	odd	s-a	even
s-a	even	c-h	odd
a-s	even	h-c	odd
m-c	odd	a-s	even
t-r	odd	m-c	even
m-c	even	t-r	odd

counterbalanced to control for such variables as: the sequence of conditions during which the letter-sounds were learned; which letter of a pair was presented first in any session; or in the first or second problem; and the assignment of letters to either of the two conditions.

Results

The data were analyzed and graphed on the basis of the daily test probes that followed the training sessions for each pair of letters. In this experiment there were six trials for each pair of letters (under each condition), three for each of the letter-sounds trained. The percent correct under each condition was based on the number of correct responses out of the total of six possible for each pair.

The mean percent correct responses for all subjects to the pretests, daily test probes and posttests are presented in Figure 1. The last two days of the training sessions when each subject met the 100% correct responding on the test probes are not included in the data.

Correct Responses on Daily Test Probes
Group Data

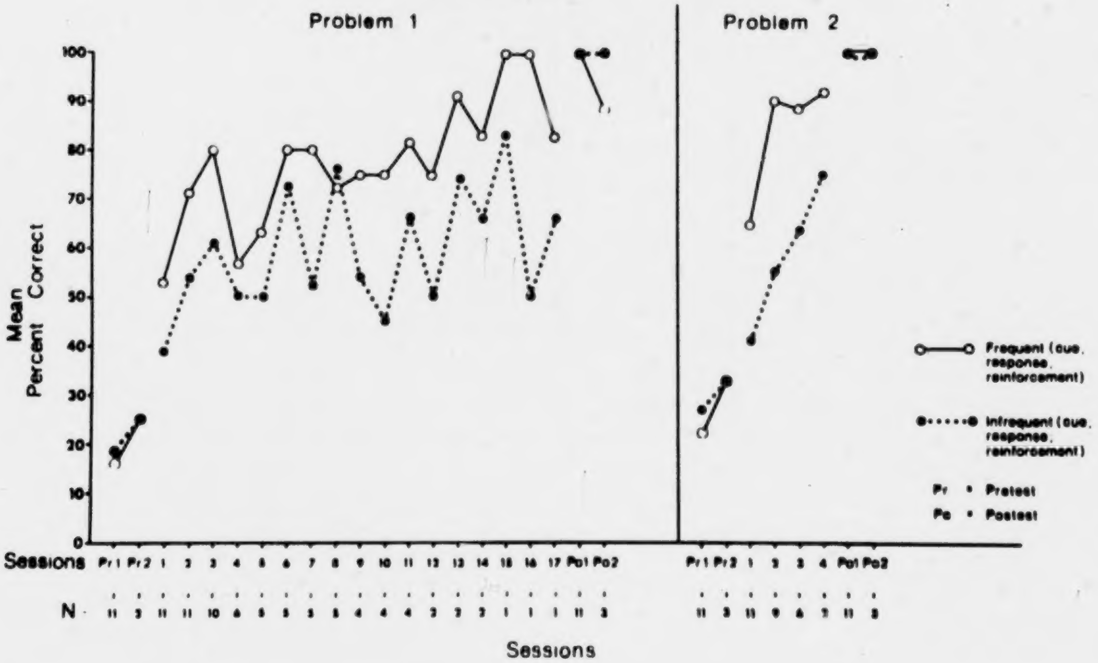


Fig. 1. Group data of mean percent correct responses on daily test probes for each session of Experiment 1 under both conditions for Problems 1 and 2.

The data from the pretests indicate a correct responding level of about 20% for the letter-sounds assigned to both the frequent and the infrequent conditions. As training progressed, the frequent condition (solid curve) produced a greater percent correct responding than the infrequent. The only exception is Session 8 of Problem 1. The same trend also occurred in the replication (Problem 2), even though fewer sessions were necessary for the children to reach criterion for all letter-sounds.

The top graph of Figure 2 shows the mean number of sessions to criterion performance for the frequent and infrequent conditions for the group.

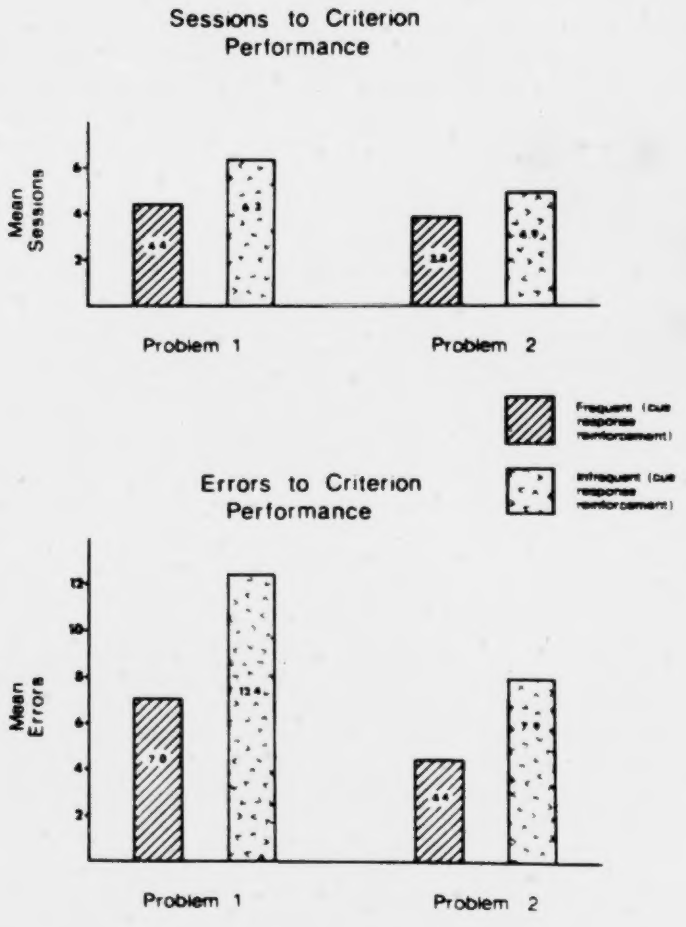


Fig. 2. Group mean number of sessions to criterion performance under both conditions of Experiment I for Problems I and 2. Group mean number of errors to criterion performance under both conditions of Experiment I for Problems I and 2.

The frequent condition (striped bar) had a mean of 4.4 sessions to criterion performance as compared with 6.3 in the infrequent condition (spotted bar). The same trend occurred in the replication (Problem 2).

The bottom graph of Figure 2 shows that 7.0 mean errors to criterion occurred during the frequent condition as compared with 12.4 during the infrequent condition. Again this trend occurred in replication.

Examination of the individual data showed that all subjects tended to have similar individual graphs to the group graph. These results indicate that even though the amount of instructional information was held constant, letter-sound acquisition occurred more rapidly; that is, in fewer sessions and with fewer errors during frequent response and reinforcement than during infrequent response and reinforcement conditions. However, it is not clear from these data whether the more frequent subject-responses or the more frequent reinforcement was the major variable contributing to the effect. Therefore, two control studies were run using a similar individual analysis comparison with group controls, as described in Experiment I.

CONTROL STUDY A

In this study, four preschool children served as the subjects. The number of reinforcers given to each subject varied, but all other training procedures remained constant. This meant that the information given to the child, the experimenter cues (instructions) and the number of responses were the same. This study was carried out as the first control procedure where the effects of frequent and infrequent reinforcement were studied under the frequent procedures used in Experiment I.

Training Sessions

Each child was trained on two pairs of letter-sounds for each problem. One pair of letter-sounds was taught with ten reinforcers per session and the other with four. In this study there were always ten subject responses under both conditions. The training and test probe procedures were the same as in Experiment I, except that some subject responses were not followed by a token or social reinforcement due to the infrequent reinforcement condition.

Design

Each subject was his own control for the main independent variable in that each child learned two pairs of letter-sounds during the same sessions, but one pair was acquired under a condition of frequent reinforcement, the other under a condition of infrequent reinforcement. Group controls were incorporated as described in Experiment I.

Results

The results of this first Control Study are shown in Figure 3. The mean percent correct for all four children during frequent reinforcement (solid line) was not appreciably different from that during the infrequent reinforcement training (dotted line) in that both curves overlap across sessions. This same observation of no differences for frequent or infrequent reinforcement also held during the replication.

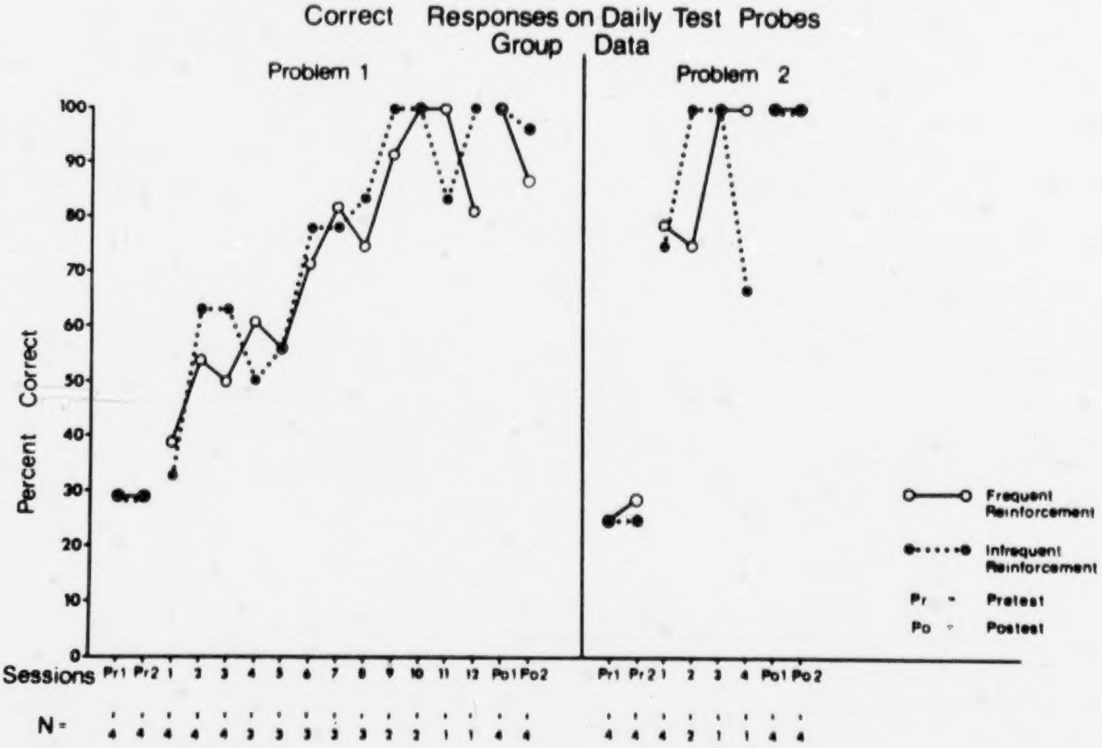


Fig. 3. Group data of mean percent correct responses on daily test probes for each session of Control Study A under both conditions for Problems 1 and 2.

Figure 4 shows the group mean number of sessions to criterion (top graph) and errors to criterion (bottom graph). These data were also similar during frequent and infrequent reinforcement, 7.8 and 7.0 sessions to criterion (respectively), and 16.5 and 15.5 errors to criterion. Replication (Problem 2) resulted in an equal number of sessions (1.5) and errors (2.0) to criterion during both frequent and infrequent reinforcement conditions.

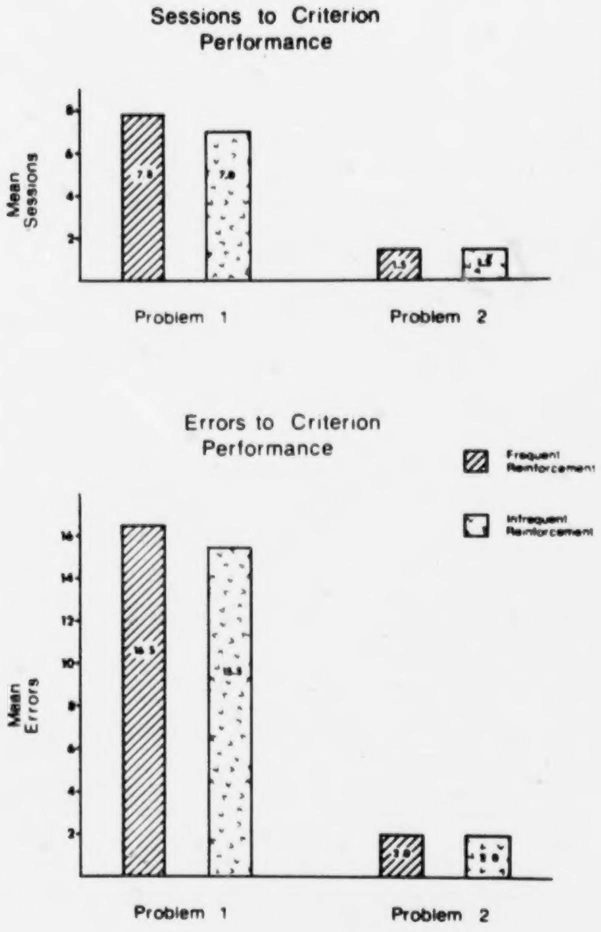


Fig. 4. Group mean number of sessions to criterion performance under both conditions of Control Study A for Problems I and 2. Group mean numbers of errors to criterion performance under both conditions of Control Study A for Problems I and 2.

Examination of the individual data showed that the four subjects tended to have similar individual graphs to the group graph. None tended to show any individual bias toward one method.

CONTROL STUDY B

In this study, with two preschool children as subjects, the number of experimenter cues and subject-responses were varied but the other training procedures remained constant. This meant that the information given to the child and the number of reinforcements were the same. This study was carried out as a second control procedure where the effects of frequent and infrequent experimenter-cues and subject-responses under the same training procedures of Experiment I were studied.

Training Sessions

Each child was trained on two pairs of letter-sounds in each session, one pair with ten responses, the other pair with four. There were always ten reinforcers delivered during the training of both pairs of letter-sounds even though the child responded only four times during the infrequent response training condition. This was done by giving a token and social praise for attending to the stimulus letters at the time when the child would have responded if he were in the frequent response conditions. In all other respects the training procedures were the same as described in Experiment I.

Design

Each subject was his own control for the main independent variable in that each child learned two pairs of letter-sounds during the same session but one pair was acquired under a condition of frequent responding, the other under infrequent responding.

Results

The mean percent of correct responses on the daily test probes for both subjects is shown in Figure 5. Higher percent correct test probes resulted during the frequent response condition (solid line) than during the infrequent response condition (dotted line) except for Sessions 8, 10 and 11. This was also true during the replication.

The group mean number of sessions and of errors to criterion are shown in Figure 6 and indicates faster acquisition of letter-sounds trained during the frequent response condition (7.5 sessions) than during the infrequent response condition (8.5 sessions). The difference in acquisition was even more evident in mean number of errors (bottom graph). The frequent response condition had an average of 17 errors as compared to 26 errors during infrequent training. The same trend occurred during the replication.

A summary of the results of Experiment I and the two Control Studies demonstrated that children learn more rapidly in that they take fewer sessions and make fewer errors during conditions of a high ratio of required responses than during a low ratio of the same. In addition, it appeared that frequent responding was a more critical variable than frequent reinforcement for faster acquisition.

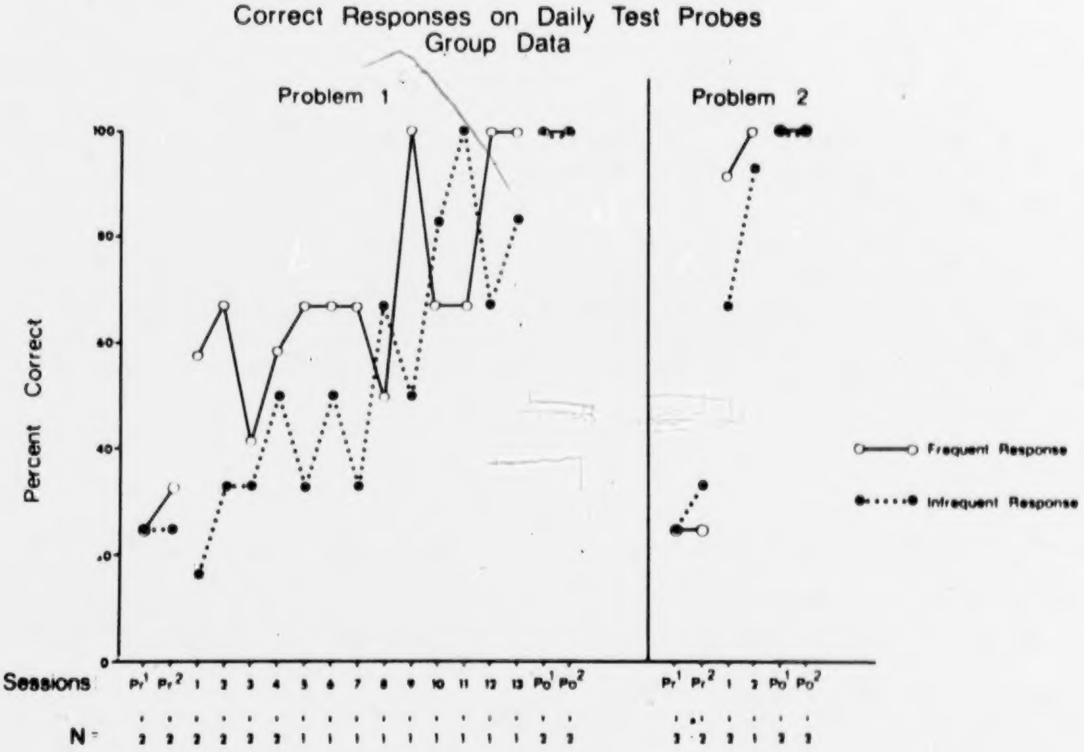


Fig. 5. Group data of mean percent correct responses on daily test probes under both conditions of Control Study B for Problems 1 and 2.

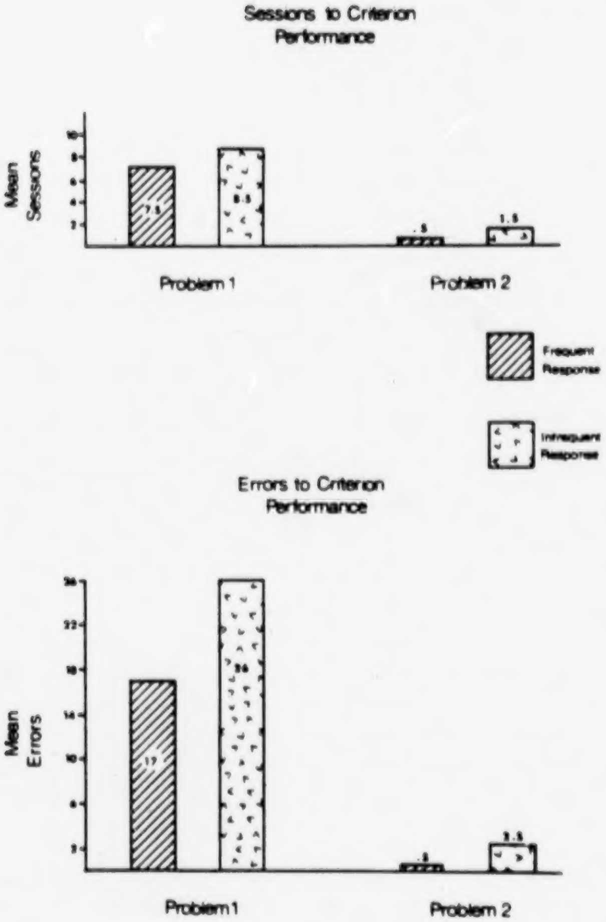


Fig. 6. Group mean number of sessions to criterion performance under both conditions of Control Study B for Problems 1 and 2. Group mean number of errors to criterion performance under both conditions of Control Study B for Problems 1 and 2.

EXPERIMENT II

Following the development of a teaching procedure and the investigation of some of the variables contributing to letter-sound acquisition by that procedure, the effects of a teaching material were then studied on letter-sound acquisition. In this experiment the subject used a tactile response of moving the finger over sandpaper and painted letters during training on letter-sounds. This study, in which eight preschoolers participated, was undertaken to demonstrate the type of "materials research" that should be carried out and a design to use to help minimize individual differences.

METHOD

Materials

The materials used during the training sessions and test probes were the Montessori sandpaper letters. The pairs of letters trained were *m-c*, *r-h* and *b-f*. The latter set of letters was a substitute pair. Duplicate letters were painted to match the color of the sandpaper letters. The other distractors, *x* and *z*, used only during the test probe, were cut from cardboard to match the color of the sandpaper and paint.

Pretest and Posttest

The children were given pre- and posttests of pointing to the sandpaper letters, following the same procedures and with the same materials used in Experiment I.

Training Sessions

On the first session of training the experimenter sat by the side of the subject and demonstrated how to follow the outline of the letters with the index finger. Then during the rest of the training only the subject used a tactile response of moving the finger over the sandpaper and painted letters when instructed to, "Feel (the letter) and say (the sound of the letter)."

Each letter was presented with five informative statements such as, "This is the first sound you make when you say *boy*." Each statement was then followed by a request for a response to feel and say, such as, "Feel *bŭ* and say *bŭ*." The other letter was then taught in the same manner. For each subject, one letter was assigned to the sandpaper condition and the other letter to the paint condition. The first presentation of each session alternated between sandpaper and paint. Following the training of sandpaper and paint letters, daily test probes on those two letters were administered, where the subjects only pointed. The six-item probes contained three trials for each letter-sound taught. Each trial involved the two letters, plus one of the cardboard distractors, *x* or *z*.

Training was continued until the children reached a criterion of two consecutive days of 100% correct responding plus 100% on the posttest.

Design

Experiment II employed both an individual analysis design and group control as used in Experiment I. Each subject was his own control for the main independent variable under study. Also, each subject engaged in a subsequent replication of the first condition, but with two new sounds.

Results

The pretest, daily test probes and posttest data for sandpaper and paint conditions are presented in Figure 7. Each data point represents the mean percent correct for that session for the group and does not include the two successive days of 100% correct responding prior to the posttest. The data for the pretests (on the lower left corner) indicate a similar correct responding level for the letters assigned to both the sandpaper (solid line) and paint conditions (dotted line). As training progressed the sandpaper condition always resulted in higher percent correct responses except for Session 11 when they were the same. Similar results occurred in the replication.

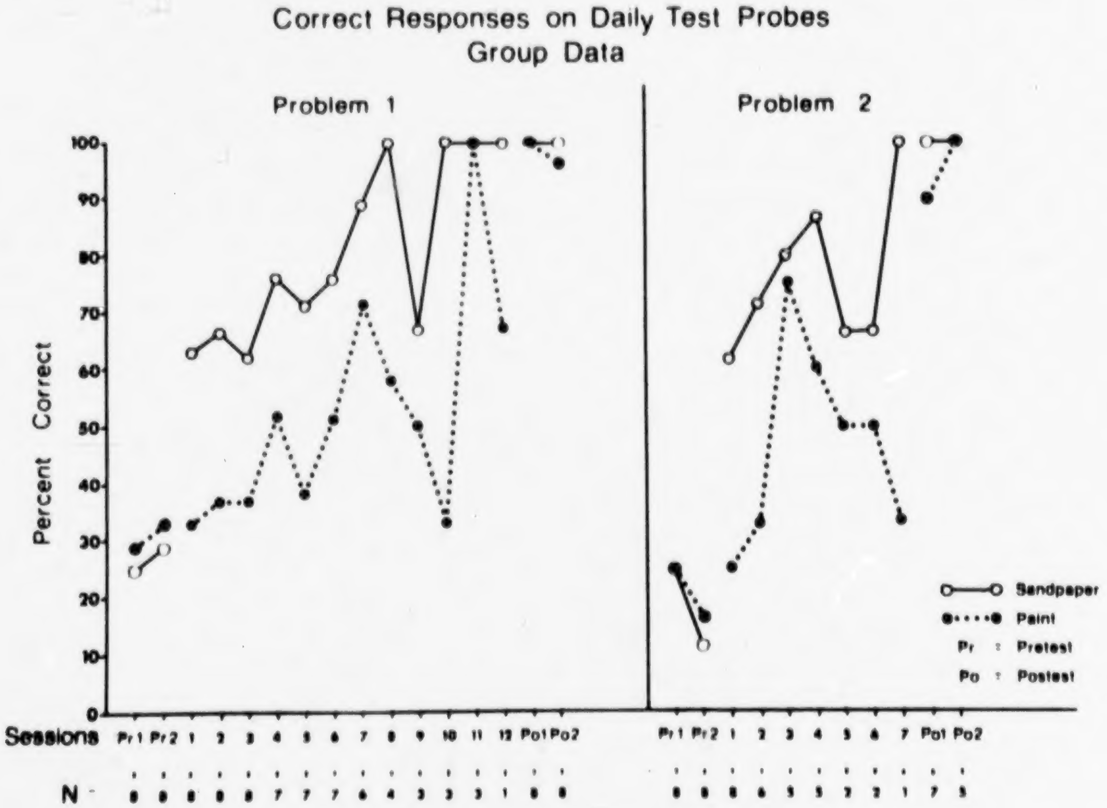


Fig. 7. Group data of mean percent of correct responses on daily test probes under both conditions of Experiment II for Problems I and 2.

In the top graph of Figure 8, the mean sessions to criterion for the sandpaper condition (hatched bar) was smaller than the mean sessions to criterion for the paint condition (white bar), 4.8 and 7.5, respectively. This was also true in the replication (2.3 and 3.6).

The mean errors to criterion during the paint condition (11.8) was more than twice the number of errors to criterion during the sandpaper condition (5.6). Errors to criterion during the replication were 6.0 and 2.8, respectively.

These results indicate that tactile responding to sandpaper letters results in faster acquisition with fewer errors and sessions to criterion than tactile responding to painted letters.

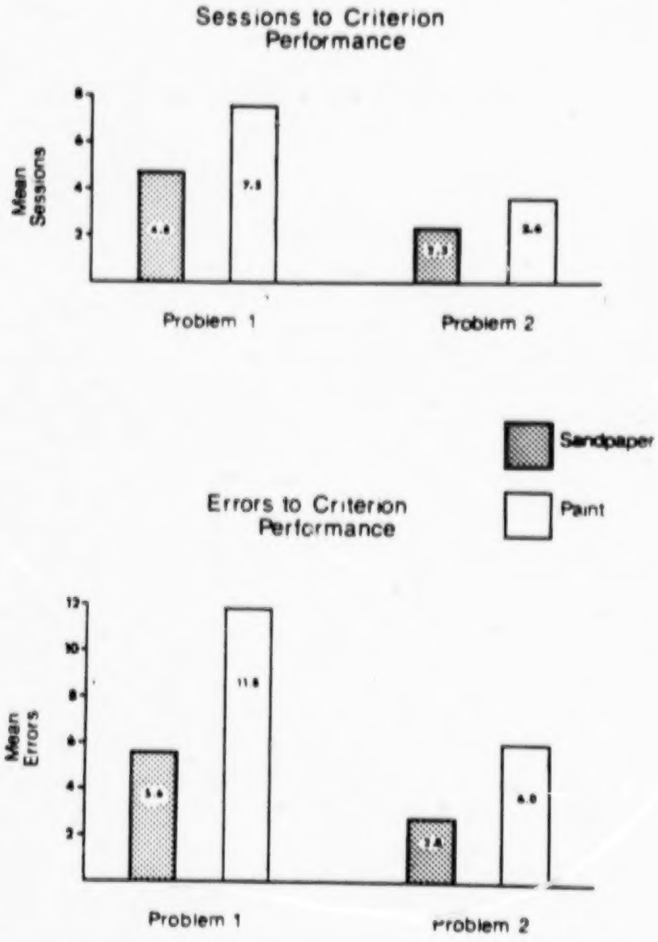


Fig. 8. Group mean number of sessions to criterion performance under both conditions of Experiment II for Problems I and 2. Group mean number of errors to criterion performance under both conditions of Experiment II for Problems I and 2.

DISCUSSION

The results of Experiment I with the Control Studies suggest that teachers should be using procedures that incorporate frequent child responses. Training in behavior modification techniques most often emphasizes the use of reinforcement in teaching. However, these results indicate that *frequent responding* during acquisition should be, at least, equally emphasized. Also, lengthy instructions or short "lectures" before asking children to respond are probably not as efficient as teaching procedures to increase child responding. The technique of obtaining frequent responses from children either in individual or group settings needs to be explored to its limits.

Many reasons for why a higher ratio of responding resulted in more rapid learning with fewer errors could be given. The most obvious explanation would be a "physiological" link. That is, more sensory receptors and effectors are involved in frequent responding. The difference between responding to letters during frequent and infrequent conditions in this study was not simply a ratio of five to two. The children looked at the stimuli, heard the sound, verbalized the sound and pointed to the letter many more times in the frequent than in the infrequent conditions. Thus, a compound effect may have occurred with the more frequent sensory receptors (hear and see) and effectors (say and point). This could have contributed to more accurate responding and more rapid acquisition.

It is important to note that the subjects who required the largest number of training sessions did not readily respond to instructions. Their correct responses during acquisition fluctuated over sessions. The behaviors of these subjects indicate that before acquisition of a concept can occur some children may require a shaping procedure to first acquire attending behavior.

In addition to teaching procedures, teaching materials may have differential effects on skill acquisition. Experiment II was designed to incorporate a teaching procedure already found to be superior to another and then vary a material to see if further differences could be obtained. The results suggest that teachers should explore novel or different-from-usual teaching materials. It is difficult to argue that sandpaper is the only material that could produce more rapid acquisition. Perhaps sandpaper is novel in the child's stimulus environment, whereas painted stimuli are more common. It could also be argued that anything different (e.g., velvet or fur) would have the same effect. On the other hand, it could be an interaction of both stimulus novelty and increased tactile stimulation which produced faster learning.

Using each child as his own control with respect to the main independent variable is essential in studies (like those presented here) which seek optimal procedures for effecting learning acquisition. Each child's replication of his own first study by engaging in a second study (Problem 2) increased the reliability and generality of the findings. Although there were always fewer sessions and errors during acquisition of the second problem, the differences obtained during any one of the studies on the second problem were consistent with those under the first problem. Only with this type of an individual analysis design are differences in children's acquisition rates controlled.

For example, Figures 9 and 10 show the data of Experiment I if they were treated in the manner of the traditional group design. Eight of the subjects were taught the same pairs of letter-sounds (*r* and *t*, and *b* and *n*). Since the conditions under which the pairs were taught were counterbalanced, then one half the subjects could comprise a group treated with frequent conditions, the other half, a group treated with infrequent conditions. If the decision of which subjects were to receive frequent and infrequent training were decided by a coin flip, the research design would become a traditional group design.

As Figure 9 shows, the group treated with frequent reinforcement and responding (solid line) did not acquire the letter-sounds more rapidly than the group experiencing infrequent presentation of these variables.

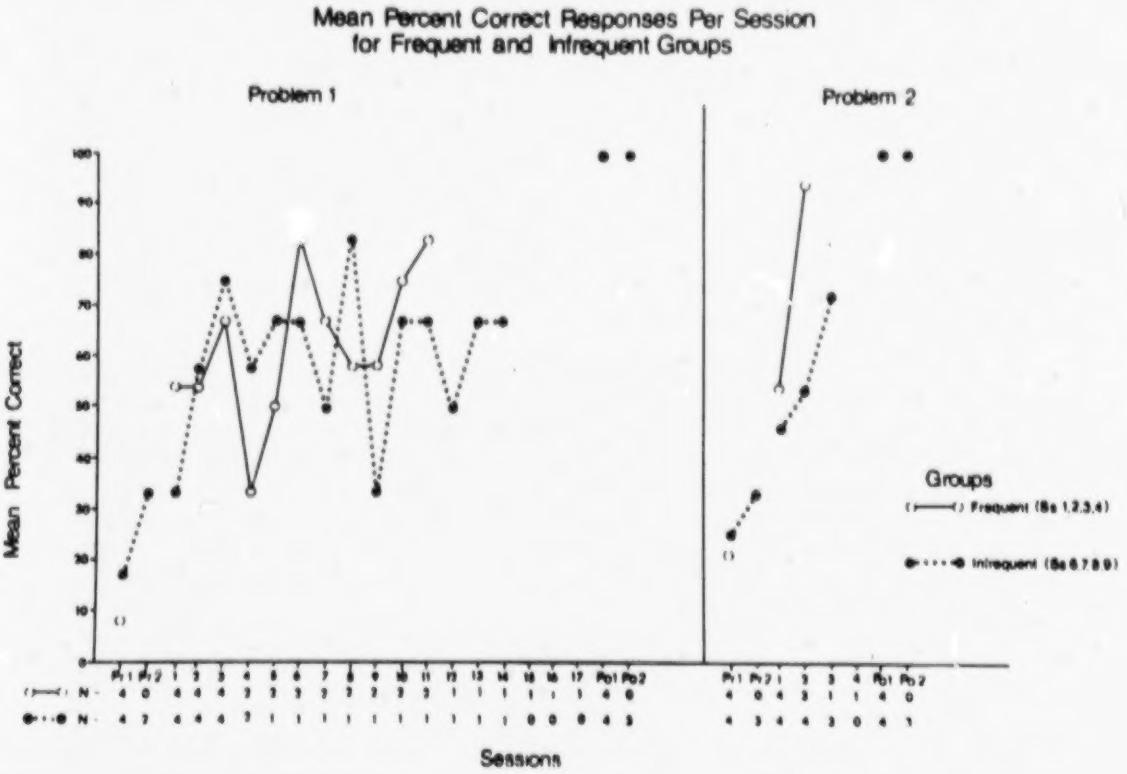
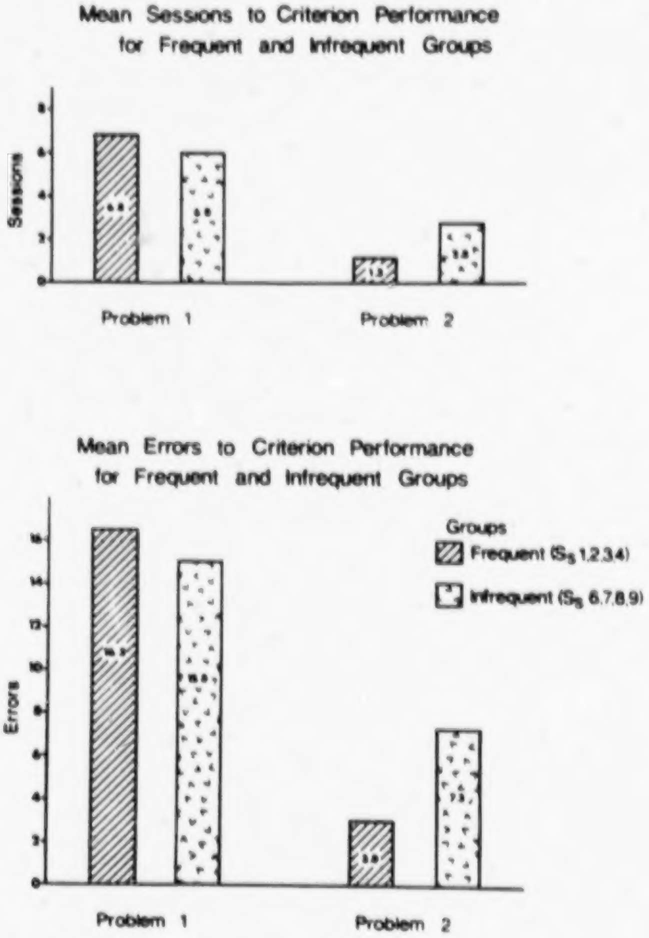


Fig. 9. Mean percent of correct responses per session for frequent and infrequent groups for Problems I and 2.

Further, the upper graph of Figure 10 shows that the frequent group (striped bar) required a slightly higher mean number of sessions to criterion (6.8) than did the infrequent group (6.0, spotted bar). The bottom graph of Figure 10 shows that the subjects acquiring letter-sounds in the frequent condition also had a higher mean error score (16.3, striped bar) than those learning in the infrequent condition (15.0, spotted bar). In this design, a statistical analysis


Mean Sessions to Criterion Performance for Frequent and Infrequent Groups

Mean Errors to Criterion Performance for Frequent and Infrequent Groups

Fig. 10. Mean number of sessions to criterion performance under both conditions of Experiment I for frequent and infrequent groups for Problems I and 2.

Mean number of errors to criterion performance under both conditions of Experiment I for frequent and infrequent groups for Problems I and 2.

would probably result in "no significant differences" and educators would be caught in the same dilemma that has plagued them for years.

These latter data did not support the conclusions we derived from the individual analysis design. However, the replication results of Problem 2 are consistent with our original results. The differences resulting from the two analyses of the same data can be attributed to individual differences in acquisition rates. Because of these individual differences, utilizing the individual as his own control in two different treatment conditions is obviously preferable to comparing treatments between groups of different children. When individual learning rates are controlled by design rather than by statistics, the formation of the laws of psychology and education will advance more rapidly.

REFERENCES

Blank, M. and Frank, S. M. Story recall in kindergarten children: effect of method of presentation on psycholinquistic performance. *Child Development,* 1971, 42, 299-312.

Gliner, C. R. Tactual discrimination thresholds for shape and texture in young children. *Journal of Experimental Child Psychology,* 1967, 5, 536-547.

Hagen, J. W. and Kingsley, P. R. Labeling effects in short-term memory. *Child Development,* 1968, 39 (1), 113-121.

Lovitt, Thomas C. and Curtiss, Karen A. Effects of manipulating an antecedent event on mathematics response rate. *Journal of Applied Behavior Analysis,* 1968, 1 (4), 329-335.

Weir, M. W. and Stevenson, H. W. The effect of verbalization in children's learning as a function of chronological age. *Child Development,* March, 1959, 30 (1), 143-151.

OPEN CLASSROOMS: SUPPORTERS OF APPLIED BEHAVIOR ANALYSIS[1]

mary mc neil; susan hasazi
adler muller; martha knight[2,3]

During the current school year, the entire teaching/administrative staff of the Hinesburg Elementary School in Hinesburg, Vermont is working with consulting teachers to develop and implement individualized instructional programs for children in regular classrooms with deficit behaviors. The consulting teacher and the staff work together to provide this special education through the following steps (McKenzie, 1971):

1➔ The classroom teacher identifies the child in need of special education and refers him to the consulting teacher.

2➔ The classroom teacher and the consulting teacher define behavioral objectives and devise measurement procedures that will reflect achievement of the defined objectives.

---

[1] This research was supported by a grant (USOE-OEG-0-71-1637) from the Bureau for the Education of the Handicapped, U.S. Office of Education; by a grant (55414-69) provided by Title VIA, Elementary and Secondary Education Act from the U.S. Office of Education, Department of Health, Education, and Welfare, to the Vermont State Department of Education; and by the Chittenden South Supervisory School District, Shelburne, Vermont.

[2] Mary McNeil and Susan Hasazi are currently Consulting Teachers in the Chittenden South Supervisory School District, Shelburne, Vermont; Adler Muller is the principal of the Hinesburg Elementary School in Hinesburg, Vermont; and Martha Knight is an instructor in the Special Education Program at the University of Vermont, Burlington, Vermont.

[3] The authors are indebted to teachers Patricia Coleman, Patricia Flood, Ruth Hesslink, Kathleen Kashuba, Judith LaForge, Mary Pree, Lori Stabb and Charles Wilson for their precision in implementing the individualized learning programs.

3➡ The classroom teacher obtains daily reliable measures of the child's achievement.

4➡ The consulting teacher determines that the child is eligible for special education if the daily measures reflect achievement below the defined behavioral objectives.

5➡ The classroom teacher and the consulting teacher inform the parents that their child is eligible for special education and obtain parental permission to provide special education.

6➡ The consulting teacher develops an individualized learning program for the child.

7➡ The classroom teacher implements the individualized learning program and obtains daily reliable measures of achievement.

8➡ The classroom teacher and the consulting teacher continue to work with the child until he has achieved the defined behavioral objectives and is therefore no longer eligible for special education.

The Hinesburg School presented a unique challenge to the consulting teacher because it represented a philosophy of education traditionally considered antagonistic to a behavioral approach such as that used by the consulting teacher. This philosophy is based on the *open classroom* model exemplified in the British Primary Schools. In the open classroom model, large groups of children are provided with a variety of activity centers through which they may move freely and converse as they work (Stroud, 1970). Typically, work projects are initiated and carried out by individuals or small groups. This flexible arrangement affords maximum interaction between children not possible in the traditional self-contained classroom.

The *open classroom* approach was new to the Hinesburg School. During the summer months both the principal and teachers had agreed to remove partitions between the previously self-contained classrooms in order to provide large open spaced areas for each group. Two and three member teaching teams were assigned from 48 to 72 children homogeneously grouped according to age.

CLASSROOM MANAGEMENT

Referral

The first referral for special education in the open classroom did not involve the typical *one* child and *one* teacher, but 64 children and three teachers. The teachers reported that until organization and management problems were remediated individual behavioral deficits could not be identified accurately.

The teachers stated their major problems as being related to time spent in moving from one activity to another, individualizing instruction, pupil motivation, and working effectively using a team teaching approach.

Measures

The dependent variable measured was minutes when the entire group was involved in the following non-academic tasks:

a▶ taking roll call for attendance.
b▶ recording requests for milk and hot lunch.
c▶ distributing work materials.
d▶ moving to and from work areas.

Procedure

The class was divided into thirds and each group was instructed as to which of the three entrances was to be used by that group. As each child entered the classroom he was required to check his attendance and request for milk and hot lunch on a chart devised by the consulting teacher. This procedure was specifically designed to reduce group time spent in non-academic tasks.

Prescription packets were compiled daily for each child. These packets contained a prescription sheet listing the child's assignments as well as the areas of the room and the time for working on each assignment. These were helpful to the teacher in determining exactly where each child was in academic areas. Also contained in the packet were the previous day's corrected papers and *fun sheets*. This procedure was designed to reduce time spent in distributing work materials and moving to and from work areas.

A point system was established which allowed groups of five children to earn free time at the end of each academic block. The teacher directed one child in each group to record points for the entire group when they emitted the following behaviors:

Behavior	*No. of Points*
●a) moving from one activity center to another within two minutes.	②
●b) no talking above a whisper when moving.	①
●c) having necessary materials ready *(pencils, paper, books, worksheets)*.	①
●d) face oriented towards teacher, materials, or peers.	①
●e) raising hand for teacher.	①
●f) waiting quietly for teacher.	①

If at least ten points were earned during the specified academic activity time the children were allotted five minutes of free time to select a game or continue to work on their packets. Groups with fewer than ten points remained working at the subject area with their teacher. This procedure was primarily designed to increase individual involvement in academic tasks.

Results

Figure 1 represents a record of group time spent in non-academic activities during a two-hour period on consecutive school days. During baseline conditions minutes of non-academic time ranged from 25 to 31 minutes. As you can see, this is about one fourth the time spent in non-academics. During the experimental condition when prescriptive packets, points and free time were instituted, minutes of non-academic time dropped from four to nine minutes. This level was maintained during subsequent postchecks taken at bi-weekly intervals. The consulting teacher aide recorded minutes of non-academic time on one occasion during each condition.

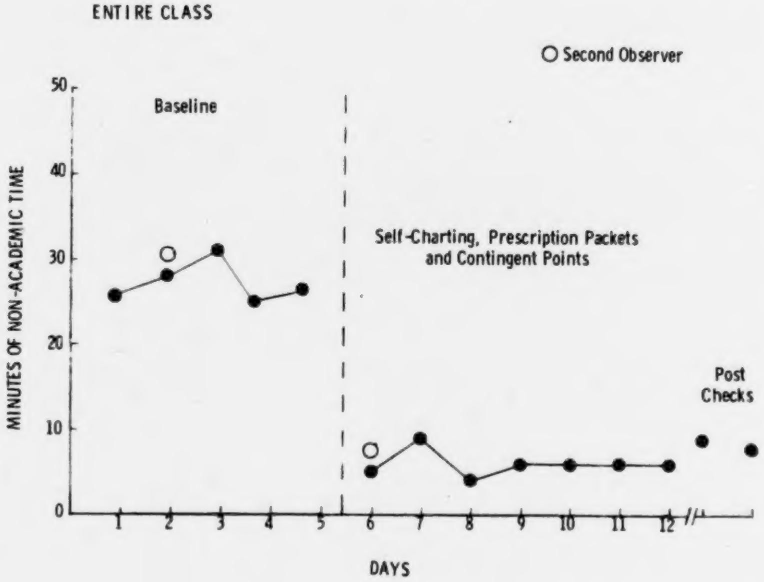


Fig. 1. Minutes of group involvement in non-academic tasks during two-hour sessions on consecutive school days.

DISCUSSION

The *Open Space Environment* facilitated frequent observation of the referred class by other teachers in the building. These teachers were anxious to develop similar procedures in their areas and requested our services.

Once the overall management problems were remediated, the teachers could then identify children with behavioral deficits. Thus, as a result of the work in the referred classroom, teachers referred thirty individual children for consulting teacher services within the next few weeks.

Since this time, points have been removed and prescriptive packets are written together with the child on a weekly and contractual basis. It is no longer necessary to use free time as a reinforcer since children have become increasingly more involved in choosing and successfully carrying out their own programs.

✪IMPROVING MATH PERFORMANCE OF AN EIGHT-YEAR-OLD GIRL✪

Referral

Barbara, an eight-year-old girl, was one of 45 children in a third year open classroom. Her teachers described her as a highly emotional child, with severely deficit academic skills. Initially, Barbara was referred because of incomplete assignments and slow progress in mathematics.

Measures

Daily measures consisted of computing the percent correct of Barbara's daily mathematics assignments. The teacher recorded the data daily.

To insure reliability of these measures a second observer occasionally scored the same assignments on at least one occasion during each condition.

Procedure

After Barbara received her daily mathematics assignment the teacher monitored the completion of the first three problems. The first problem was done entirely by the teacher in Barbara's presence. The second problem was done cooperatively by both the teacher and Barbara. The third was done by Barbara in the teacher's presence. Barbara then continued by herself until the paper was completed. Upon completion, the teacher corrected the assignment immediately. If the paper was between 90% and 100% correct, Barbara would take the paper and a card to the principal's office for a private conference. If he was not in his office the index card was left on his desk. This was a signal to the principal to visit Barbara in her classroom for their conference.

Results

During baseline, when the procedure was not employed, the percent correct of mathematics responses ranged from 0% to 100%. When conferences were initiated, the correct increased to optimum levels. Agreement between observers was 100%.

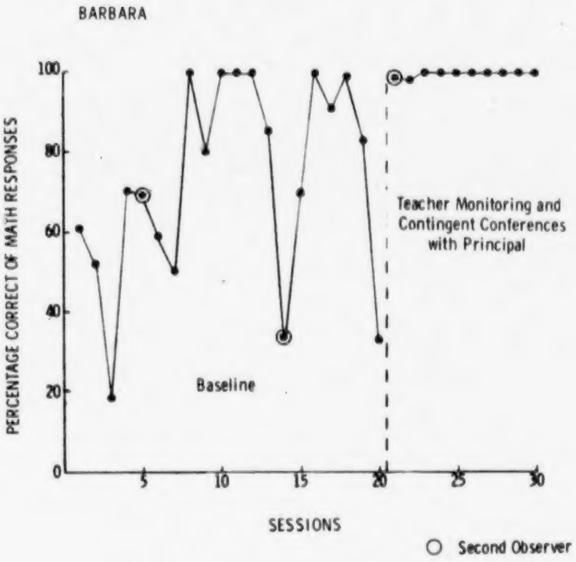


Fig. 2. Percentages of correct math responses during consecutive daily sessions.

DISCUSSION

The teachers reported that Barbara now seems eager to do her work well. She has asked if she might have a similar program for reading. The teachers are most eager to expand such a program for the entire day.

REDUCING VERBAL AGGRESSIONS OF A 4TH GRADE BOY

Referral

Alan, a nine-year-old boy, was one of 62 children in a fourth year open spaced classroom. He was referred because of his reported verbal aggressions toward other children.

Measures

Verbal aggression was defined as any loud abusive statement such as "Shut up," or "I'm going to get you after class," directed to another child. The number of verbal aggressions was tallied by the teacher during this half hour session each day. At the conclusion of the daily session the total number of aggressions was recorded.

To insure reliability of these measures, the consulting teacher occasionally served as a second observer on at least one occasion during each condition.

Procedure

Alan was given a data recording sheet similar to that used by his teacher. It was explained that he and his teacher would be charting his daily behavior during a specified time period. A criterion of no more than two instances of verbal aggressions per session was set. At the completion of the half hour period, Alan and the teacher compared data. If the number of tallies did not exceed two on both the teacher's and Alan's data sheet a card was sent home stating that Alan had been a cooperative student that day.

Results

During BASELINE 1 and BASELINE 2, when the procedure was not employed, the number of verbal aggressions ranged from two to seven during the daily half hour session. During *Contingency 1* and *Contingency 2* when Alan began recording his own data and receiving a card to be taken home, the number of verbal aggressions ranged from zero to two. Agreement between observers was 100%

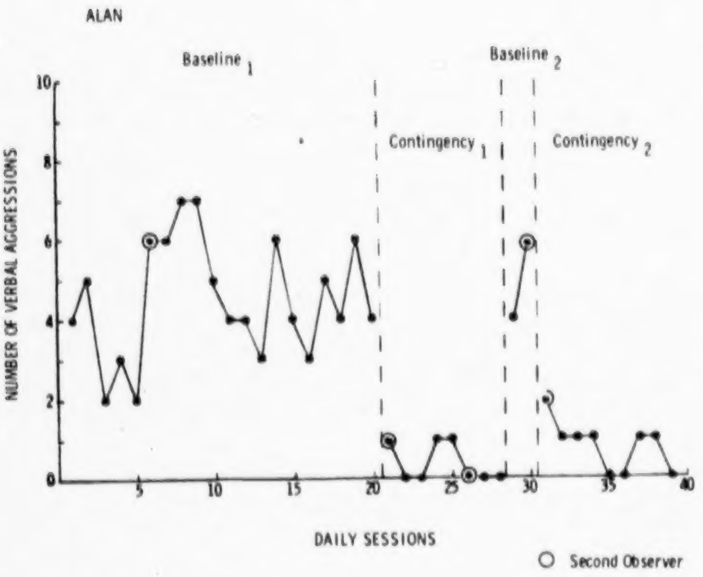


Fig. 3. Number of verbal aggressions during consecutive daily 30-minute sessions.

DISCUSSION

The teachers were so pleased by the results that they decided to use this procedure throughout the entire day. They have divided the day into two periods: *morning* and *afternoon*. Alan is currently taking data for the entire day and can earn one card for each period. Additionally, the teachers have decided to lower the criterion to one instance of verbal aggression per period. They hope to reduce the criterion to zero in the near future.

INCREASING PARTICIPATION, WORK COMPLETION, AND ACCURACY BEHAVIOR IN A GIFTED THIRD GRADER

Referral

Kathleen, an eight-year-old girl, was one of 45 children in a third year open space classroom. She was referred because she had recently transferred to the school and was not performing at her reported potential.

Measures

Daily measures included the percentage of assigned tasks completed during the entire school day, as well as the percentage of completed and correct math responses during a daily 30-minute session.

To insure reliability of these measures the consulting teacher occasionally served as a second observer on at least one occasion during each condition.

Procedure

An individualized prescription packet was designed for Kathleen which contained a schedule of tasks as well as work materials for the day. Difficult tasks, such as mathematics, were scheduled to be completed first during a 30-minute period. An index card which had the numeral corresponding to the number of assignments for that day was given to Kathleen. After she completed each task on the prescription sheet she raised her hand and her teacher checked to see if the work was completed. If the total number of points equalled the numeral on the card she earned the card.

The teacher then wrote a message to the parents such as, "Kathleen has completed all of her work today." This card was then given to Kathleen to take home to her parents who had agreed to place the card on the refrigerator and to spend extra time with her expressing their satisfaction with her performance at school.

Results

During BASELINE 1 and BASELINE 2, when the procedure was not employed, the measured behaviors were low. During *Contingency 1* and *Contingency 2*, when prescription packets and teacher/parent praise were in effect, the measured behaviors rose dramatically. During *Contingency 2* the behaviors were maintained at optimum levels. Agreement between observers was 100%.

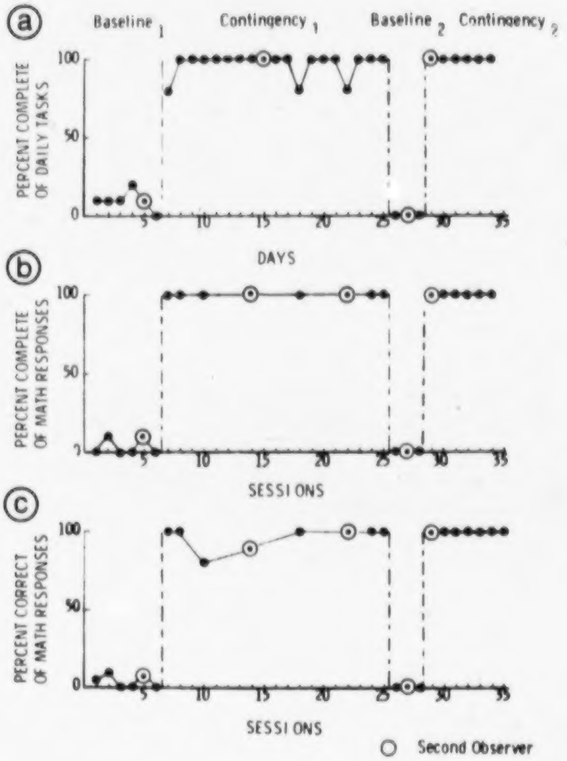


Fig. 4a. Percentages of complete tasks during consecutive days.

Fig. 4b. Percentages of complete math responses during consecutive daily sessions.

Fig. 4c. Percentages of correct math responses during consecutive daily sessions.

DISCUSSION

The teachers have reported that Kathleen is now progressing at an accelerated rate in all of her subjects and that this type of structure is beneficial for children in an open classroom.

ELIMINATION OF BELLIGERENT AND TRUANT BEHAVIOR IN TWO PRE-DELINQUENT BOYS

Referral

The subjects were two boys in a sixth year open space classroom, in which there were 69 children and three teachers.

George was referred because of failure to do his assignments and refusal to comply with teacher requests. For the past five years George had been living with foster parents.

Jack was referred as a result of his truant behavior and his general attitude of non-compliance. He had been referred on several occasions to juvenile authorities for attempted arson and theft.

Both boys were several years below grade level in academic subjects.

Measures

Participating behavior was defined as face oriented toward teacher and/or assigned materials. This behavior was observed at three minute time samples during the same half hour period each day. Days on which the boys completed assignments were also recorded.

To insure reliability of these measures the consulting teacher occasionally served as a second observer on at least one occasion during each condition.

Procedures

Individualized prescription packets were designed for the boys. They included a schedule for the day and assignments from such materials as *Sullivan Programmed Math and Reading*, *The New Practice Readers*, *Reading for Comprehension* and individually planned units in social studies and science. Activities such as recess, music, lunch and physical education were also on the daily prescription sheets.

Upon completion of each assignment the teachers initialed a card in the prescription packet. If the entire packet was completed the initialed card was sent home. Two completed cards earned a conference with a consulting teacher. Six completed cards earned a trip of their choice with a consulting teacher. Trips earned thus far include a canoe trip, hockey game, movies, lunch at McDonald's and a visit to the police station.

Results

During BASELINE 1 and BASELINE 2, when the procedure was not employed, Jack's percentage of participating time ranged from 0% to 100% and George's also ranged from 0% to 100% per half hour session while task completion was variable. During *Contingency 1* and *Contingency 2*, the boys' percentage of participating time maintained at 100% and tasks were completed consistently.

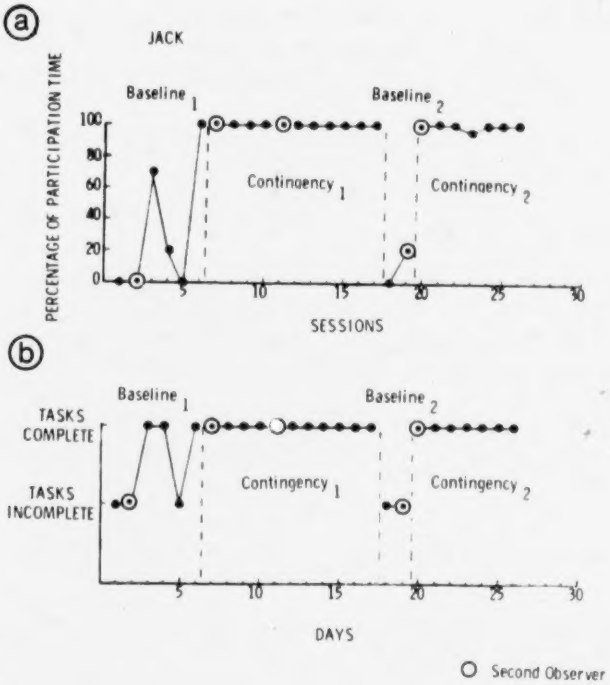
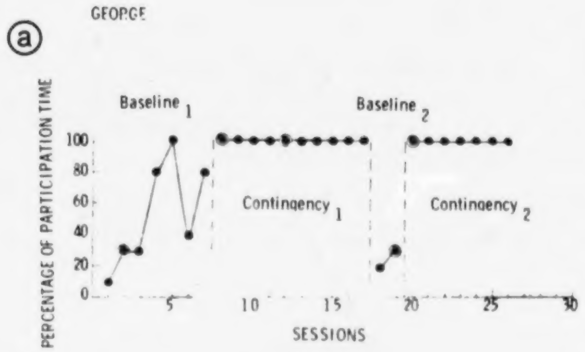


Fig. 5a. Percentages of participation time during daily 30-minute sessions.

Fig. 5b. Tasks complete or incomplete during consecutive days.



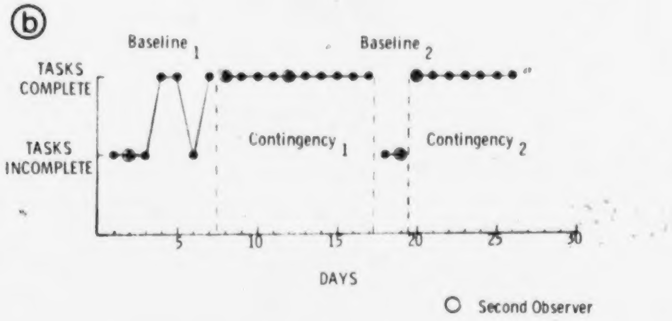


Fig. 6a. Percentages of participation time during daily 30-minute sessions.

Fig. 6b. Tasks complete or incomplete during consecutive days.

DISCUSSION

Teachers and administrators have reported that both boys are now *model* students. They have suggested expanding the program to include behaviors such as smiles, conversing with adults and sharing with peers.

A measured decrease in the number of the boys' verbal refusals was also reported, but there was no measure of the number of teacher requests that would occasion verbal refusals.

Currently the schedule of reinforcement is being thinned.

REFERENCES

McKenzie, H. S. Special education and consulting teachers. In Clark, F., Evans, D. and Hammerlynk, L. (Eds.), *Implementing Behavioral Programs for Schools*. Champaign, Illinois: Research Press Company, 1972, 103-125.

Stroud, Marion. *Perspectives on The Prospect School*. The Prospect School, North Bennington, Vermont, 1970.

INTEGRATION OF NORMAL & HANDICAPPED CHILDREN IN A BEHAVIOR MODIFICATION PRESCHOOL: A CASE STUDY[1]

k. eileen allen
paulette m. benning
w. thomas drummond

Recent research at the Experimental Education Unit (EEU) Model Preschool includes a pilot study of the effects of a 50-50 enrollment mixture of 16 "normal" and handicapped children in a preschool class. Procedures for studying this and related questions are still gross, although we have published some tentative findings (Allen, Turner, and Everett, 1970). Nevertheless, we are still "feeling our way" and have few definitive answers about our eventual methodology. One of our first efforts is to collect comprehensive data on a number of normal and atypical children, using a data system that enables us to look at the "whole child." We think we are beginning to get some clues. For example, in the case study that follows, the data indicate that mere integration in a group composed of 50% normal peers did not eliminate all of the emotionally disturbed child's maladaptive behaviors. Specific intervention procedures were required for several behaviors while several other behaviors showed spontaneous improvement. Because we are interested, too, in ascertaining how much data-taking a supervising preschool teacher can manage, we cite this case study because it also speaks to this question.

SUBJECT AND SETTING

Julie was three years and eleven months old when she was enrolled in the Model Preschool. She was an unusually small child for her age though there were no untoward medical signs except that of arrested hydrocephalus in early infancy. The medical follow-ups indicated no evidence of subsequent physical or neurological impairment. Developmentally, she exhibited poor large motor skills and infantile speech patterns. In addition, she displayed an extensive repertoire of inappropriate and maladaptive social behaviors toward children and adults. With children, her most conspicuous responses were avoidance or attack.

---

[1] This paper was presented at the Third Annual Conference on Behavior Analysis in Education at the University of Kansas in Lawrence, Kansas, May, 1972. The Model Preschool Center for Handicapped Children is funded in part through P. L. 89-230 Title VI, Part C, Sec. 623. Dr. Alice H. Hayden is the Project Director; Dr. Norris G. Haring is the Co-Director.

With adults, she was totally noncompliant. She screamed "No" to any request, however mild; any insistence by an adult (parent or teacher) provoked a full-fledged tantrum in which she threw herself face down on the floor, kicking and screaming hysterically.

Her previous school placement had been a preschool for emotionally disturbed children in the psychiatric department of a children's hospital. The staff there felt that Julie had made some progress but that she might progress more rapidly if she could interact with a more normal peer group than was available in the hospital preschool setting. Two members of the EEU preschool staff observed Julie in the psychiatric setting and they agreed to enroll her in a preschool class at the EEU. In this class were 16 children, half of whom were classified as normal, the other half as handicapped. The handicapping conditions varied: emotional disturbance, mental retardation, orthopedic involvement, language impairment. Classroom personnel were a head teacher, an assistant teacher, and an intern advancing a master's degree in special education. Several other adults were in the classroom each quarter--students from various disciplines for whom the preschool staff provided graduate and undergraduate training.

PROCEDURES

The basic Washington Social Code (Bijou, Peterson, Harris, Allen, and Johnston, 1968), with a number of variations, was employed as the data system. Data were recorded at 10 second intervals for 20 to 70 minutes during free-play indoors. Length of sample time varied with the amount of personnel on hand each day. Specifically, data were collected on:

(1) quality of social behavior, indicating whether the child was engaged in cooperative, parallel, or isolate play;

(2) amount of verbal behavior with peers;

(3) amount of time not engaged in purposeful activities (e.g., wandering, watching, rocking or thumbsucking, apart from a play activity or material);

(4) amount of disruptive behaviors such as screaming, having a tantrum, running off with other children's possessions;

(5) number of attacks (kicks, hits, or bites) that made physical contact with another child; and

(6) amount of teacher attention, defined as any interval in which a teacher touched, spoke to, or presented Julie with material of any kind.

Six days of baseline data were collected; on the fifth day a new category of maladaptive behaviors was added: throwing and dumping play materials and equipment.

At the end of six days, the head teacher declined to continue baseline, arguing convincingly that Julie's steadily increasing disruptive behaviors and ever more ferocious physical attacks were interfering drastically with the other 15 children's school program. Several of these children were three-year-olds adapting to their first group experience. Experimentalists engaged in classroom studies know how imperative it is for successful research to have the teaching staff agree to the experimental procedures. It is sometimes necessary to be less rigorous than one would like in order to maintain teacher cooperation. Thus, a compromise was reached and *Phase I* commenced.

Phase I

The intervention tactics used in *Phase I* were a compromise, as mentioned above, between the experimenter's wish to achieve a more stable baseline and the teachers' insistence that something had to be done. After examining several possibilities the teachers agreed to operate negatively on only one behavior--the one they felt was most traumatizing to the other children--Julie's physical attacks. She was to be put in time-out for each attack. They contrived a time-out spot by placing two sets of heavy lockers in a partially closed "V" position. When the child was put in time-out, the "V" was pushed to a fully closed position, providing complete isolation within the classroom setting. Two more categories of data were added: amount of time Julie spent in time-out and amount of tantrum time during time-out. Another revision was made in the experimental procedures: the teacher collecting the data acted as the time-out agent to insure that every physical assault received an immediate negative consequence. This teacher also dispensed positive reinforcement according to routine procedures in effect for all children--intermittent social reinforcement of appropriate behaviors. (In Julie's case, bare approximations to appropriate behaviors received positive reinforcement.) All other disruptive behaviors--dumping, throwing, snatching, screaming and having tantrums--were ignored so far as possible. A series of language sampling observations was also scheduled at this time in order to assess potential qualitative improvement in Julie's verbal repertoire.

Phase II

The second set of intervention procedures, labeled "activity decision required," was aimed at reducing Julie's random wanderings about the room. The staff agreed that she most often engaged in disruptive behavior or physical assault during her aimless cruising. These procedures were used: each time Julie left an activity she was required to state where she was going to play next. Usually the teacher offered two or three suggestions: "There is room for you at the easel, or you can go to the puzzle table, or play with the blocks." If Julie refused to state a choice she was led to the locker area, but not put in time-out. The teacher stood with her back to Julie, preventing her return to the main play area until she indicated verbally where she would play. Such decisions were heartily reinforced by the teacher who then accompanied Julie to the area of her choice and helped her to get started in play. Every effort was made during this period to teach Julie how to use play materials, for she had not even rudimentary skills with blocks, manipulative toys, creative materials, or dramatic play activities. Time-out for physical aggression was still in effect during *Phase II*.

Phase III

Julie's throwing and dumping of play materials continued to be a major disruption in the classroom. One reason, of course, was that while teachers could do a heroic job of ignoring her, it was almost impossible to prevent peers from reinforcing Julie with negative verbalizations when she swept their half-finished puzzles off the table, dumped the contents of their paint cans on the floor or on them, or poured water over their sleeping dolls. In fact, argued the teachers, Julie was becoming so aversive to all the children that it would take until Doomsday for her to be accepted as a member of that peer group unless further steps were taken immediately to reduce her obnoxious behaviors. Thus *Phase III*--time-out for all throwing and dumping episodes in addition to *Phase I* and *Phase II* conditions:

(1)▶ timing-out physical attacks;

(2)▶ ignoring screaming and tantrums;

(3)▶ requiring a play-choice decision;

(4)▶ intermittently reinforcing approximations to appropriate play; and

(5)▶ teaching use of play materials.

Phase IV

Because the children continued to shun Julie, a fourth phase, structuring cooperative play activities was instituted to insure that Julie would be included. For example, certain children were designated to play with her for part of the free-play period in a particular area: "John and Julie and Sara and Shelli are to play in the housekeeping corner." The rationale for this phase was the expectation that children who were temporarily legislated to play with the semi-reformed Julie would discover that she had shed many of her aversive properties, and that teachers thus would have the opportunity to reinforce Julie for appropriate play with peers and to reinforce the children for playing cooperatively with Julie.

RESULTS

Baseline Period

During baseline, Julie's physical attacks on children rose to a high of 68 during one 60-minute observation period (Figure 1), while throwing and dumping episodes reached a peak of 74 (Figure 2A). Disruptive behaviors rose to 48% (Figure 2B); the amount of time not engaged in purposeful activities climbed steadily to 88% (Figure 3). Amount of verbalizing to peers was negligible, the high being 6% (Figure 4A). Cooperative interaction with peers reached 30%, but was declining steadily, and amount of time spent alone was as high as 77% (Figure 4B). Teacher attention to Julie averaged 15% during the baseline period (Figure 5).

Phase I (time-out for physical attacks)

Time-out for physical attacks resulted in an immediate and drastic decrease in attacks. Over the five-day *Phase I* period, the range was two to seven episodes (Figure 1). Concurrently, throwing and dumping episodes decelerated to a low of five episodes in two observation periods (Figure 2A). Julie's disruptive behaviors also decreased markedly to a low of four episodes (Figure 2B). (Neither of these behaviors, it must be remembered, was under treatment during *Phase I*.) Amount of time not engaged in purposeful activities dropped to 34% on one day (Figure 3); however, much of Julie's time was spent in time-out during this period, so that she had much less time available to her for "cruising." Verbalizations to peers increased slightly to 12% (Figure 4A); cooperative interaction with peers remained at approximately the same average, 25%, while isolate play decreased (Figure 4). Again, this decrease was largely due to the time Julie was spending in time-out--she simply had less time at her disposal for isolate activities. Teacher attention to Julie in *Phase I* increased slightly to an average of 24% (Figure 5) mainly because of the number of times teachers took her to the time-out area; the amount of time Julie spent in time-out ranged from 7% to 30% per *Phase I* session (Figure 6).

Phase II (activity decision required)

Even though *Phase I* produced a decrease in several of Julie's undesirable behaviors, the teachers were concerned about how much time she spent cruising about aimlessly when not in time-out. Therefore, *Phase II* was initiated, requiring Julie to state where she would play next and keeping her hemmed in a corner of the room until she did so. Teachers also continued the *Phase I* conditions of time-out for physical attacks, ignoring all disruptive episodes and intermittent positive social reinforcement for appropriate behavior. As can be seen in Figure 1, physical attacks dropped to a near zero rate and stayed there.

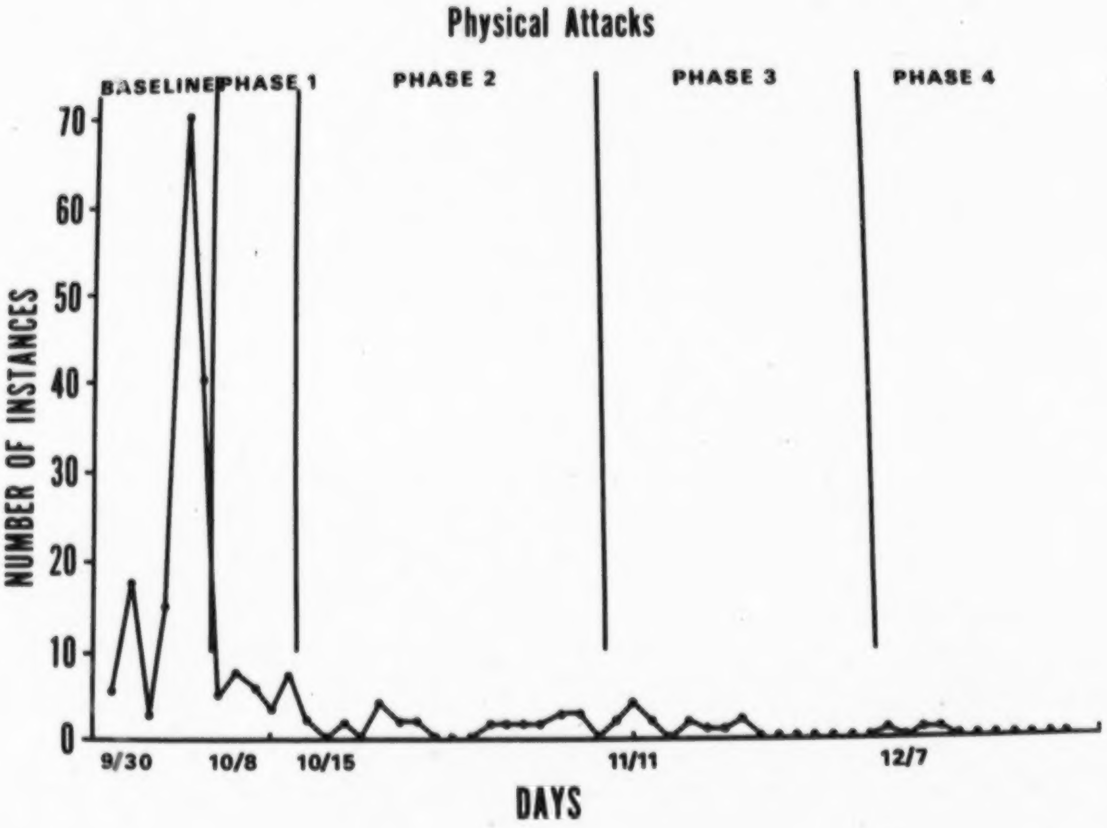


Fig. 1. PHYSICAL ATTACKS. *Phase IV* data are collapsed into 5-day units. The median data point of each unit is plotted.

However, Julie's throwing and dumping episodes rose sharply, reaching a peak of 56 episodes on one day (Figure 2A). Disruptive episodes continued at roughly the *Phase I* low rate (Figure 2B). During the first nine days of *Phase II* percentage of time not spent in purposeful activities was very high, from 50% to 85% (Figure 3). None of this was cruising time, however; it was time Julie spent in the corner deciding where to play next. Over the next nine days of *Phase II*, the amount of time not in activities varied markedly, although on four days she spent 10% or less of her time not engaged in activities (Figure 3). Since she was spending so much of her time in a corner of the room during these first nine days, verbal behavior and cooperation with peers dropped sharply but both showed good increases during the latter half

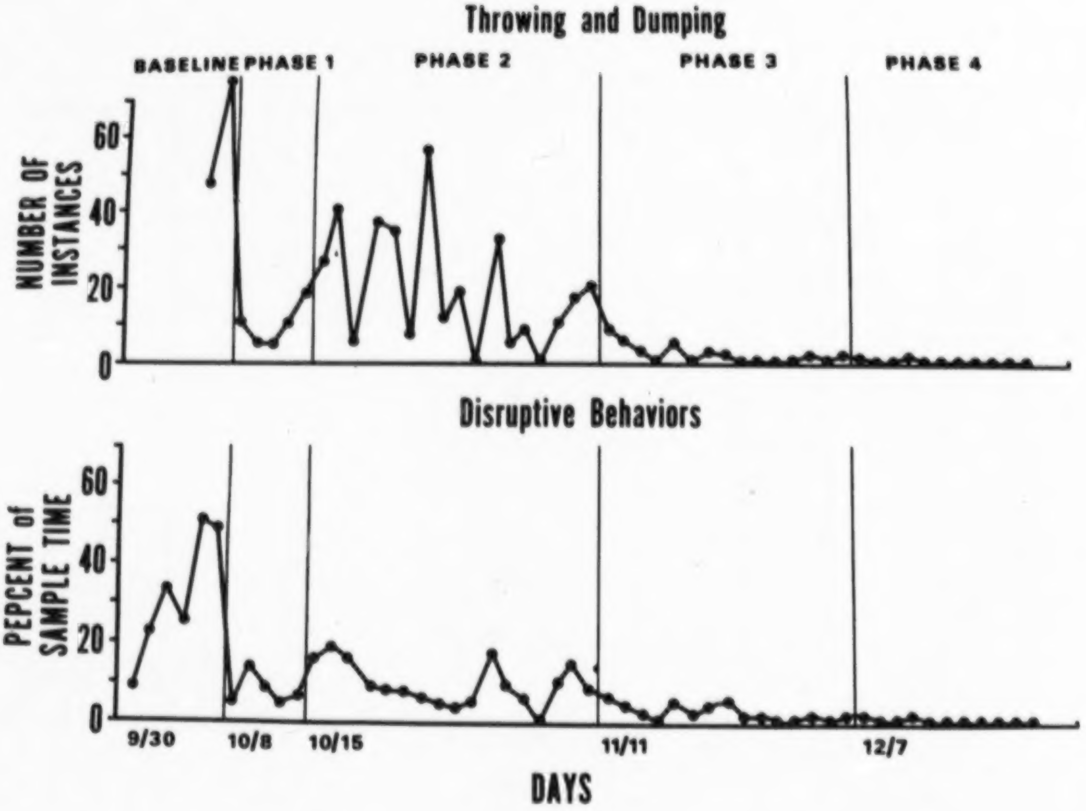


Fig. 2. A, Upper Graph. THROWING AND DUMPING.
B, Lower Graph. DISRUPTIVE BEHAVIORS.
Phase IV data are collapsed into 5-day units.
The median data point of each unit is plotted.

of *Phase II*--especially verbalization to peers: it rose to 25% on one day (Figure 4A and 4B). Teachers did not give Julie any attention when she was hemmed into the corner except for periodic cues to her to tell them where she would next play; therefore, teacher attention, except for the first part of *Phase II* was only a little greater than the typical *Phase I* rate (Figure 5). Julie also spent some time in actual time-out after making a physical attack on another, but this seldom went above 10% for any day except for one or two "bad days" (Figure 6).

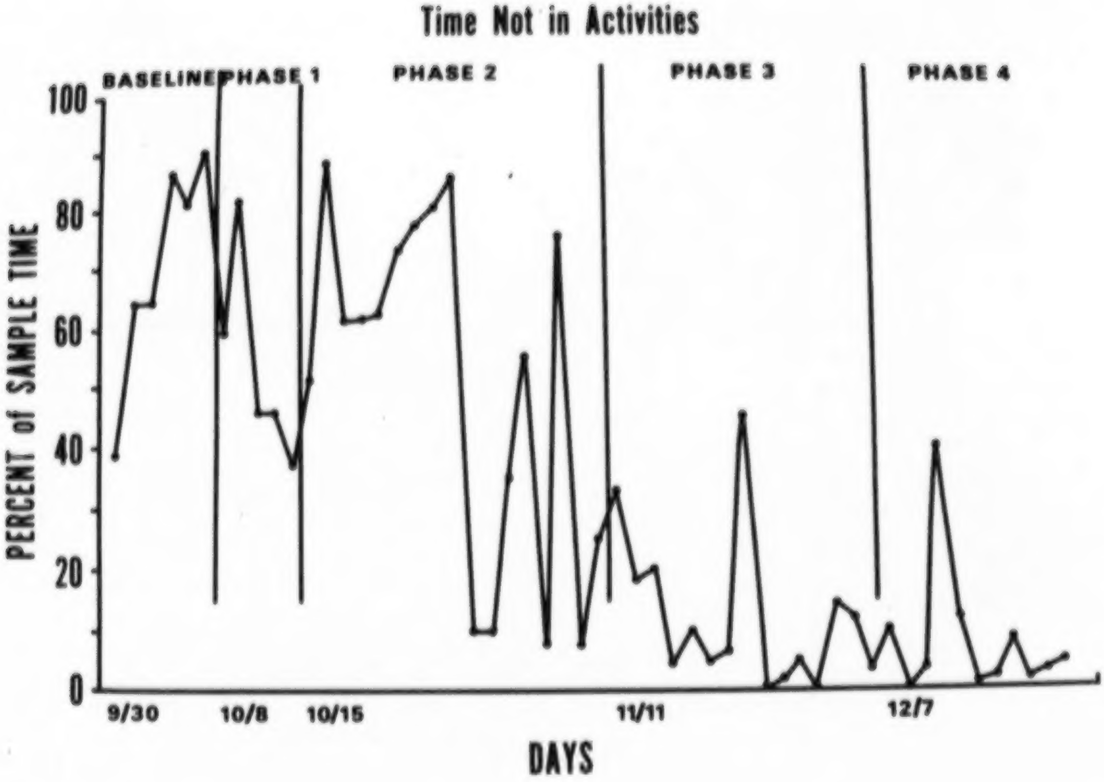


Fig. 3. TIME NOT IN ACTIVITIES. *Phase IV* data are collapsed into 5-day units. The median data point of each unit is plotted.

Phase III (time-out for throwing and dumping)

Because Julie's throwing and dumping episodes were much too frequent during *Phase II*, the teachers decided on the negative consequence, time-out, for these responses. It produced an immediate and stable decrease to a zero rate which Julie maintained for the rest of *Phase III* (Figure 2A). Physical assaults (still subject to time-out in this phase) stayed at a near zero rate (Figure 1); other disruptive behaviors also dropped to and stayed at that rate (Figure 2B). Verbalizations to peers remained at a fairly good level--from 5% to 18% per day; cooperative interaction dropped to a little less than its former level. Isolate play, however,

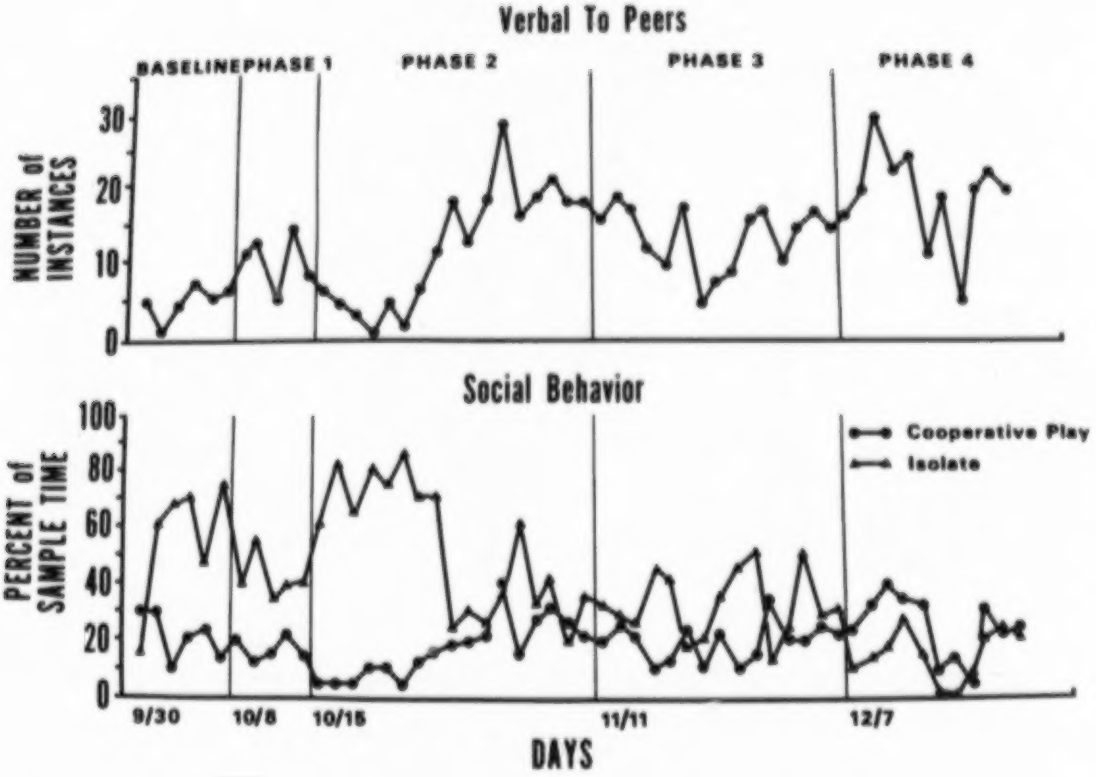


Fig. 4. A, Upper Graph. **VERBAL TO PEERS.**
B, Lower Graph. **SOCIAL BEHAVIOR.**
Phase IV data are collapsed into 5-day units.
The median data point of each unit is plotted.

decreased considerably, indicating that Julie was spending much more time near other children or in parallel play with them (Figure 4A and 4B). Teacher attention increased during *Phase II* to an average of 29% per day (Figure 5). After an initial increase in time-outs for throwing and dumping, there were no time-outs at all during the last seven days of the phase (Figure 6).

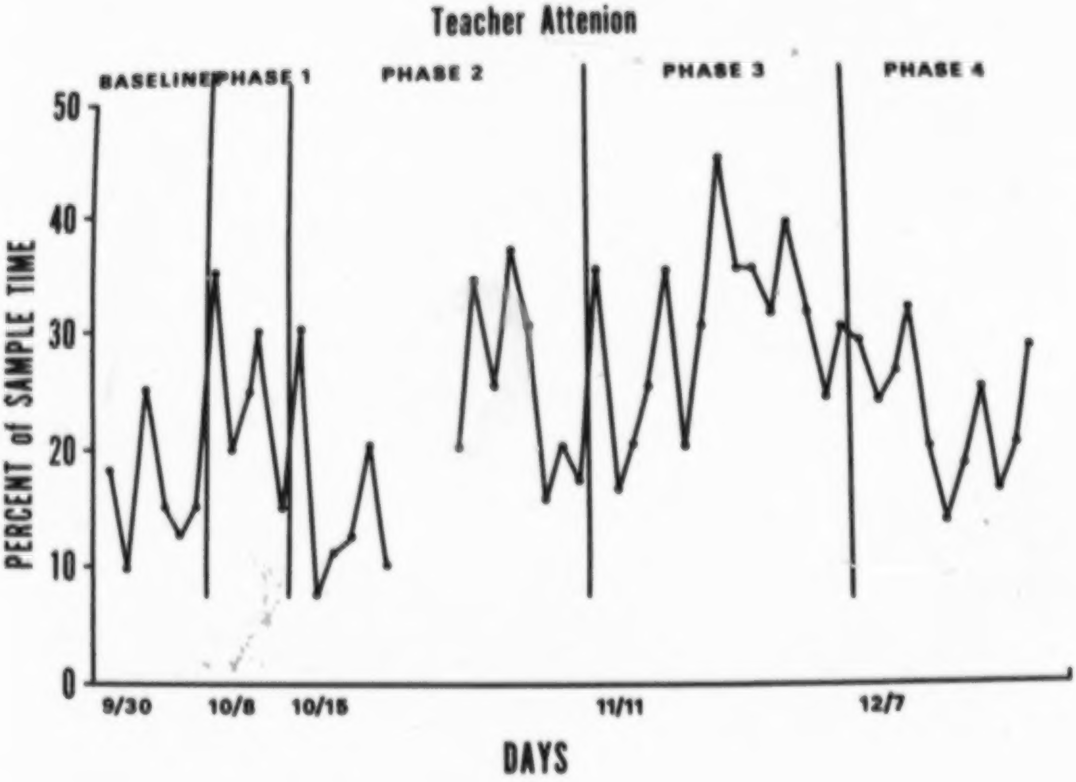


Fig. 5. TEACHER ATTENTION. *Phase IV* data are collapsed into 5-day units. The median data point of each unit is plotted.

Phase IV (structuring cooperative play)[2]

By the end of *Phase III*, Julie, according to the teachers' judgement, was not engaging in a sufficient amount of cooperative play. They began, therefore, to actually structure at least one situation per free

---

[2] *Phase IV* data are collapsed into 5-day units for more economic display; the median of each unit appears on the graphs. The expanded data are available, however.

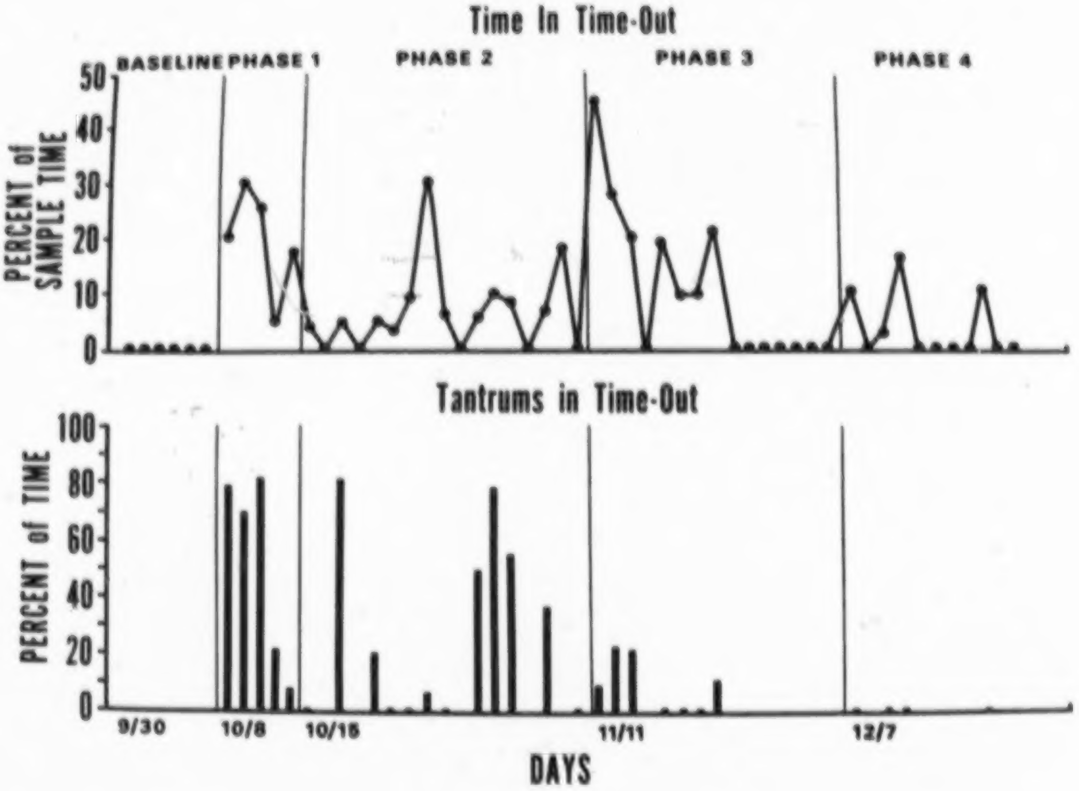


Fig. 6. A, Upper Graph. TIME IN TIME-OUT.
B, Lower Graph. TANTRUMS IN TIME-OUT.
Phase IV data are collapsed into 5-day units.
The median data point of each unit is plotted.

play session in which they instructed Julie and several other children to play together in a specific activity. They made no requirements about the length of "legislated" interaction. This program did not increase Julie's cooperative play appreciably, although on several days she did go above 25% (Figure 4B). However, the amount of isolate play decreased markedly, indicating that Julie was spending a great deal of time in proximity to children--a good first approximation to cooperative interaction.[3] Her verbalization rate also continued at an appropriate level (Figure 4A), although there was some fluctuation.

During *Phase IV*, Julie's physical attacks stayed at a near zero rate (Figure 1) as did her throwing and dumping (Figure 2A) and disruptive behaviors (Figure 2B). Time not in activities stayed well below 10% except for one day (Figure 3). In this phase, teachers discontinued the decision-making contingency. Teacher attention averaged 23% (Figure 5) although several children were sharing in this attention since cooperative play was to receive adult social reinforcement. The time Julie spent in time-out was little (Figure 6A) though it was still required occasionally. She no longer had tantrums in time-out, however (Figure 6B).

Three inter-rater reliability measures were obtained during the study, one in October, one in November, and one in January. The scores on these were 91%, 91%, and 90% respectively on total number of correctly matched response units.

DISCUSSION AND CONCLUSIONS

Though Julie's disruptive and physically aggressive behaviors were at a low level during her first few days in the classroom, these began to appear at an accelerating rate once she was thoroughly adapted to the new situation. The teachers' anxieties over the other children's well-being led to abandoning one of the experimental objectives, that of ascertaining the effect on Julie's overall behavior of simple exposure to a more normal peer group. It seems unlikely, however, that she would have improved without specific intervention; she was simply too aversive to both children and adults. Because of her aversiveness, of course, most social reinforcement from peers was directed to her undesirable behaviors. It is very difficult to get a group of preschool children to reinforce differentially a child with Julie's repertoire, especially when half of the children are themselves management problems. Furthermore, even if the teachers had been willing to program the children to do so, the extinction of Julie's maladaptive responses would probably have been too slow and arduous a process.

After taking baseline data, the teachers chose one behavior--aggressive physical attacks--as a first target behavior to modify. However, data were continuously kept on many other responses in order to ascertain what, if any, correlated changes might occur. Thus, a multiple baseline design was employed in that baseline data on certain behaviors (disruptive responses and verbal behavior to peers) were kept throughout the study while other behaviors (time not in activities, throwing and dumping and cooperative play) were removed one at a time from baseline as phase changes were made. It is interesting that while throwing and dumping and disruptive behaviors decreased during *Phase I* (time-out for physical attacks), the rate of throwing and dumping increased again within a few days though an increase of disruptive acts did *not* occur. In fact, these last responses disappeared almost entirely in the course of treatment even though they were never subjected to specific intervention.

---

[3] Amount of proximity can be obtained by combining amount of cooperative and isolate responses and subtracting from 100 for each data point.

In addition to the behaviors classified as disruptive, one other behavior, never exposed to specific intervention, Julie's verbal interaction with peers, showed marked improvement. Not only did amount of verbalizations increase spontaneously, quality also was somewhat improved. These qualitative changes are documented in a series of language samples taken on the child during the course of the study.

Another prediction, that Julie's rate of cooperative play would increase spontaneously once she became less aversive to children and had acquired some play skills, was not borne out. A specific program *(Phase IV)* was required to achieve that goal. Even then, the quantitative rate of increase was not remarkable, although the improved quality of her interaction with peers was noteworthy. These improvements are readily observable on a series of videotapes that span the study from start to finish. These tapes have been assembled into a video case study to provide additional documentation of the dramatic changes in Julie's behavior.

This study, as well as several others in progress, seems to indicate that children with severe problem behaviors are not likely to improve without specific intervention. With specific intervention, however, it appears that progress is rapid where there are normal peers to serve as models for appropriate behavior. Also, the "cost" in teacher time would seem reasonable. In Julie's case it exceeded by very little the typical amount that children receive in a well-staffed preschool. Thus, it may be assumed that other children are not being "shortchanged." Data on the normal children indicate steady progress with no deleterious side effects. Additional favorable evidence appears in their parents' responses to the program. Without exception, all parents asked for a second year's enrollment for their child, or for a younger sibling if the older child was at kindergarten age.

Finally, this study yields some tentative answers to the question, "Can teachers realistically be expected to take data, program for, and manage so diversified a group?" In this study, the head teacher collected much of the data while supervising in the classroom. Such a dual role has been carried out by teachers in other studies (Turner, Allen, and Smith, 1972, in press); therefore, data collection by the teacher is not unique. However, teacher-collected data are usually simple counts or duration measures of fairly discrete events whereas the data system for this study was much more complex. The teacher recorded 11 aspects of the child's behavioral repertoire per 10-second interval over a 20- to 70-minute observation period each day, thus making 120 to 420 recorded observations during each class session.

Another noteworthy feature of this study documented vividly on videotape is that in addition to supervisory and recording duties, the data-taking teacher also acted as the reinforcing agent for Julie. That is, the data-taking teacher dispensed much of the positive social reinforcement for Julie's appropriate behavior and served as sole agent for immediately putting the child in time-out for the designated inappropriate responses. It might be underscored that one of the reasons for assigning the recording teacher to this additional duty was to insure an immediate negative consequence. The other adults in the room, engaged in working with the other 15 children, often failed to see Julie's undesirable response or could not disengage themselves from an activity they were involved in rapidly enough to provide immediate negative feedback. There was some concern, initially, that the recording teacher's additional role of reinforcing agent would yield much lower inter-rater reliability, but this fear proved unfounded.

While it cannot be stated definitively how and in what numbers to combine normal and handicapped children in a single classroom, a series of studies such as the one described in this paper should provide a foundation for designing a long-term, comprehensive research program to ferret out the answers. Answers are needed, for we do handicapped children a severe injustice if we keep them out of the educational mainstream. Grouping atypical children together, segregating them from normal peer models, often seems to increase their atypical behaviors. Behavior analysis has demonstrated that it can help to remediate many behavior problems; surely it can be expected to yield empirical data that will help meet the urgent need to better educate handicapped children.

REFERENCES

Bijou, S. W., Peterson, R. F., Harris, F. R., Allen, K. E., and Johnston, M. S. Methodology for experimental studies of young children in natural settings. *The Psychological Record*, 1969, 19 (2), 177-210.

Allen, K. E., Turner, K. D., and Everett, P. M. A behavior modification classroom for Head Start children with problem behaviors. *Exceptional Children*, 1970, 37 (2), 119-127.

Turner, K. D., Allen, K. E., and Smith, P. K. Modification of an autistic child's verbal and hyperactive behaviors, in press, 1972.

◆◆◆◆

part 3:

PROGRAMS & TECHNIQUES IN BEHAVIOR ANALYSIS:

A Comprehensive Behavior Analysis Program for Handicapped Children

A COMPREHENSIVE BEHAVIOR ANALYSIS PROGRAM FOR HANDICAPPED CHILDREN

INTRODUCED BY: **don r. thomas**

The following set of five papers was produced in the Child Behavior Laboratory School at the University of Illinois. As a group, the papers focus on the systematic application of behavioral principles to the social and academic problems of children discarded by the public school system.

The first paper, by *Grimm* and *Bijou*, introduces the Lab School's program and provides detailed descriptions of the subject population, the school environment and the research strategy. Next, *Berner* and *Grimm* present the program used to teach oral reading and comprehension skills. In addition, this paper details the types of data used to evaluate the program and to provide information for further program development. Finally, the paper briefly presents data on the use of the reading program by a mother in her own home. The third paper in this group, written by *Rayek* and *Nesselroad*, focuses on the development of three interdependent instructional programs. The research on teaching writing, spelling, and composition is presented in terms of the analysis of tasks to be taught and the procedures which will enable the children to master those tasks. The fourth paper, by *Parsons*, covers the analysis of arithmetic behavior as manipulative operant behaviors and reviews the development of an effective arithmetic program. Finally, *Grimm* summarizes the implications of the preceding papers.

As a set of readings, the papers clearly suggest that a body of replicable techniques is under development for managing the setting events, academic stimulus materials, and teachers' behaviors which make up the child's learning environment. Clearly, psychology has something to offer education--*now!*

PRINCIPLES AND OBJECTIVES IN AN ACADEMIC PROGRAM FOR YOUNG HANDICAPPED CHILDREN[1]

jeffrey a. grimm
sidney w. bijou

One of the children enrolled in the Child Behavior Laboratory class once commented to his teacher about his difficulties, "I think I have my thinking cap on backwards." A science of human behavior that concentrates upon hypothetical processes may one day develop a cognitive hat that can be rearranged on the child's head to remediate the child's problem behaviors. However, behavior analysts have chosen not to wait and have insisted that there is no need to invent the cognitive hat. Instead, research efforts have been aimed at the arrangement of the classroom environment to promote progressive changes in the child's behavior toward desirable school objectives.

In the research to be presented, empirical behavior principles have been applied to the development of procedures and materials for the remedial teaching of young retarded and emotionally disturbed school children. A more comprehensive description of the program has been reported elsewhere (Bijou & Grimm, in press). As introduction to the following papers on the programming of cognitive school subjects, we shall describe:

1. the types of children studied;
2. the classroom situation in which the research takes place; and
3. the nature of the research.

SUBJECT POPULATION

The children, ranging in age from five to eight and coming from a variety of socio-economic classes, are referred by schools in Champaign County, Illinois. They are viewed as severe behavior problems or extreme learning disability cases neither of which the public schools feel equipped to handle. These behavior problems can be grouped into three broad categories: behavioral excesses, behavioral deficits, and inappropriate stimulus control.

---

[1] The research described here and in the papers by Grimm and Berner, Parsons, and Rayek and Nesselroad is supported by the U.S. Office of Education, Division of Research, Bureau of Education for the Handicapped, Project No. 23-3030, Grant No. OEG-0-9-232030-0762 (032).

Behavioral excesses, commonly called conduct problems, are displayed by aggressive, extremely disruptive, destructive, or hyperactive children. The child is usually *off-task*, engaging in a high rate of inappropriate, non-academic behaviors or changing activities frequently. Examples include physical or verbal threats directed at other children, destruction of books or other academic materials, and running around the classroom or climbing on furniture. Children with behavioral deficits include the withdrawn or isolate child, the child who does not talk, and the child who does not have the precurrent behaviors for academic endeavors. Examples of the latter are inability to imitate speech sounds accurately and inability to appropriately manipulate academic materials such as pencils, erasers, or pages in a book. The third category of behavioral problems displayed in the classroom, inappropriate stimulus control, refers to behaviors that occur frequently but under inappropriate circumstances. Examples are crying in response to mild frustration and subtle aversive cues, strong dependent relationships with the teacher, and inept play behavior.

Children are ordinarily enrolled in the Laboratory class for the nine month school year. In some cases, this is shortened to as little as four months and in others it is lengthened to two successive academic years. The objective for the exceptional child is to provide him with academic and social skills that enable him to return to a class in the public schools.

THE CLASSROOM SITUATION

The classroom environment includes the physical arrangement of the classroom, the structure of the class day, the behavior of the teacher and her aides, the behavior of the children as a class group, and the nature of the instructional materials and procedures. The class itself contains desks for each of the children and has a number of quiet work areas and a small playroom. Maximum class size is eight children. The teaching staff consists of one teacher, one assistant and a number of tutors. An extended description of the Laboratory class can be found in Bijou (in press).

Three classes of environmental events are systematically managed. The first class is composed of setting or contextual events. These include establishment and maintenance of a positive emotional tone in the classroom, a flexible sequence of study periods, and the physical arrangement of the classroom. If necessary, the child's initial study behavior occurs in a quiet isolated work area but eventually takes place in the noisier general classroom area.

The second class consists of programming events that are antecedent to the child's behavior and these include the instructional materials and the modeling and prompting behaviors of the teacher. In the course of the academic year the child progresses from simple to complex materials and moves from a completely tutored context to one in which a teacher merely circulates around the classroom giving relatively infrequent attention to each child.

The third class of events comprises the management of events that are consequent to the child's behavior. These include error correction procedures and reinforcers. Progressive change in the way the child's behavior interacts with consequent events is of crucial importance. Beginning with whatever will motivate the child, generalized token and social reinforcers are established. But the contrived token reinforcers are ultimately removed and even social reinforcers are partially replaced by reinforcers intrinsic to the academic materials themselves.

The classroom teacher is, of course, the most important component in the successful management of all of these events because the teacher is the person who arranges them. The teacher sets the emotional tone of the classroom. She programs antecedent events by presenting instructional materials, giving directions, and granting permissions, etc. Her mere physical presence often comes to control her pupils' study behavior. She programs consequent events by distributing her attention among the children and by correcting academic work.

One fact that is often forgotten is that the children also manage the class by controlling the teacher's behavior. A good teacher uses this relationship to improve her own teaching skills. Each time a child fails to master a lesson, he is telling the teacher that she has failed to arrange the learning environment appropriately. If the teacher is responsive to this, then it is literally true that the child teaches the teacher how to teach him.

RESEARCH

The research is both basic and applied. When designing applied research, the experimenter is controlled by the necessity of obtaining a useful end product. He seeks to discover the practical value of making specific changes in the environment. In basic research, on the other hand, the experimenter is not necessarily looking for something useful (although this may result). The basic researcher seeks to discover the behavior changes that result from specific environmental manipulations. Although the difference is one of emphasis, different research strategies result (Skinner, in press).

The basic aspects of our research deal with the identification of variables that are causally related to the development and maintenance of some desirable behavior. The research strategy begins with the definition of the behavior of interest. Baseline conditions for that behavior are then identified and stabilized. This is followed by the experimental manipulation, the recovery of baseline conditions, and a replication of the experimental treatment. The basic research is thus of the within subjects variety. Variables that are related to the development and maintenance of attending to reading and arithmetic materials have been extensively investigated. A sample of this research dealing with bringing a child's behavior under the control of addition and subtraction signs is presented in the paper on arithmetic by Parsons.

The applied aspects of the research to be reported here deal with academic program development and evaluation. After definition of the desired or terminal behaviors, the program (e.g., reading, writing, spelling, composition, or arithmetic) is constructed in relation to three things:

(1) the desired behavior;
(2) the child's current behavior; and
(3) behavior principles derived from previous applied and basic research.

Next, the program is given a try-out. If any part of it is unsuccessful, it is redesigned (i.e., a remedial unit is designed). An unsuccessful sequence is never simply repeated. Thus, program change derives from remedial units. The program change is, in turn, evaluated with other children. This means that the Laboratory's applied research is often of the between subjects variety.

BEHAVIORAL ASSESSMENT

We now turn to the problem of assessing the child's behavior in relation to his school environment. If assessment is to be useful, it should be an aid in planning, administering, and evaluating the individual child's curriculum. Standardized psychometric tests are not serviceable in this regard because they do not assess the child's environment. In addition, evaluating an academic program on the basis of standardized tests alone does not produce adequate information unless the program has been designed to teach the specific behaviors measured by the standardized tests. Thus, we have been forced to utilize (and in some cases, develop) additional assessment techniques.

The first of these techniques is the use of pre and posttests based on actual instructional materials. The pretests enable the teacher to place each child at his initial level of competence in each academic program. The use of posttests, specific to smaller subsections of each program, allows the teacher to evaluate the stability of recently acquired cognitive behavior and the effectiveness of the programs. Examples of pre and posttest assessment are included in the papers on reading by Grimm and Berner and on writing, spelling, and composition by Rayek and Nesselroad.

The second assessment technique consists of information on all correct and incorrect academic responses, antecedent events (e.g., prompts), and the contingencies for responses (i.e., the social and token reinforcers). Obtained primarily when a child is tutored, this detailed information can be used to evaluate the child's progress, the behavior of the tutor, and the specific difficulties in the academic program. Reliability with an independent observer is, of course, maintained at the 90% level throughout the year. Examples of this type of assessment data will be found in the papers on reading and arithmetic.

The third assessment technique entails the recording of the frequency of occurrence of stimulus and response events across successive time units. Frequency of occurrence data provides an objective assessment of the strength of behavior. Figure 1 represents data collected in a public classroom by an observer prior to the target child's enrollment in the Laboratory class. The child's on-task behaviors (e.g., oral reading or attending to instructional materials) and his off-task behaviors (e.g., wandering around the room or engaging in inappropriate play with another child) are coded in the top row. On-task is scored as *N* and off-task is scored as *F*. Each box represents a ten-second interval and each ten seconds is scored for the occurrence of on-task or off-task behavior. Antecedent and consequent teacher/child interactions are coded in the second row as *I*. Although the child was on-task 33% of the time, his teacher only interacted with him when he was off-task and was thus probably maintaining his off-task behavior.

Figure 2 shows the development of another child's on-task behavior in the Laboratory class. Consecutive observational sessions are plotted along the horizontal axis. The percentage of time the child was on-task in academic settings is plotted on the vertical axis. As can be seen, the child increased her on-task behavior as the year progressed.

These data simply describe the child's behavioral change. We are interested in the relationship between these changes and observable changes in the environment. One way to explore this is to indicate the onset and offset of significant functional events. Thus, in Figure 3, it can be seen that this child was off-task in the early part of her school career with and without token reinforcers. With prolonged exposure to token reinforcement, she came on-task and then remained on-task when tokens were made non-contingent.

Example of Data on One Child Taken in a Public School Classroom

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
N	N	N	N	N	N	N	F	N	F	N	N	F	N	N	N	N	N
												I					

19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
F	F	F	F	N	N	F	N	F	F	F	F	F	F	F	F	F	F
	I	I				I											

37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F

55	56	57	58	59	60	61	62	63	64	65	66						
F	F	F	F	F	F	F	F	N	N	N	N						
				I	I	I	I	I									

Fig. 1. Example of frequency of occurrence data on one child in a public school classroom.

It is also possible to obtain more detailed information on the determining conditions. In Figure 4 we have plotted the amount of time that the teacher interacted with the child when the child was on-task. This interaction steadily decreases across the major portion of the school year with token reinforcement even though the child remains on-task. Teacher/child interaction also increases when tokens are no longer used to reinforce the child.[2] Thus, the child is well on her way to becoming academically independent and the assessment data indicate to the teacher how far she has progressed toward that goal.

---

[2] In eight reliability sessions conducted across the year with this subject, reliability with an independent observer averaged 99.25% for on-task behavior 93.75% for off-task behavior, and 98.50% for the occurrence of teacher/child interactions.

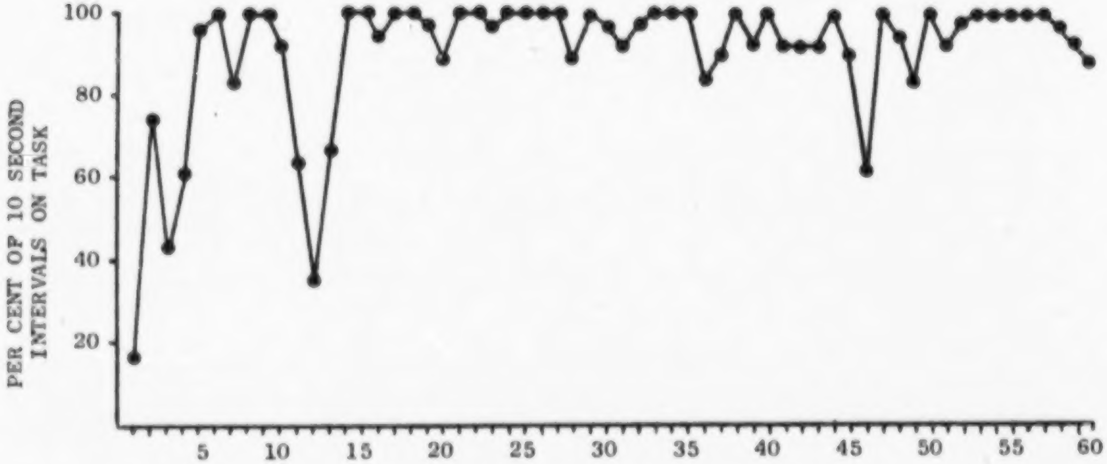


Fig. 2. The development of on-task behavior for one child.

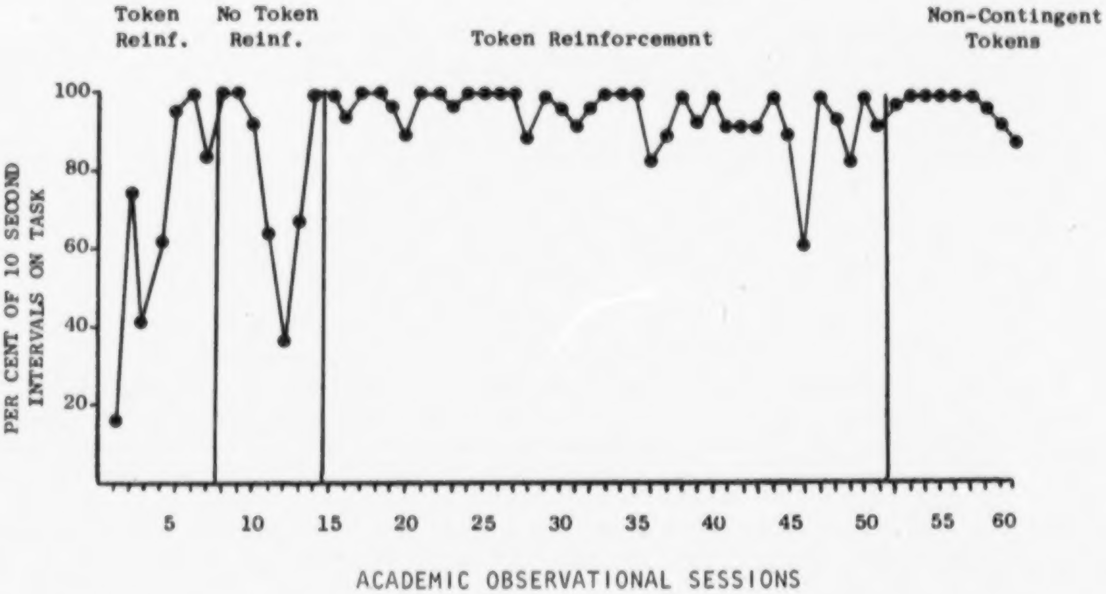
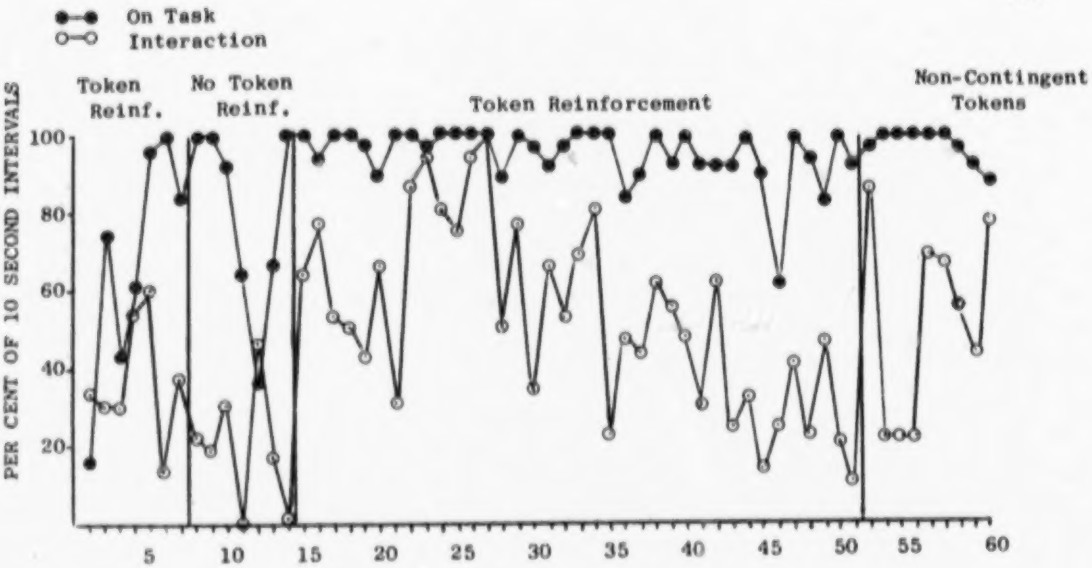


Fig. 3. The development of on-task behavior in relation to token reinforcement for one child.



ACADEMIC OBSERVATIONAL SESSIONS

Fig. 4. The development of on-task behavior in relation to token reinforcement and teacher/child interactions for one child.

SUMMARY

This paper has outlined the application of empirical behavior principles to the development of materials and procedures for the remedial teaching of young retarded and emotionally disturbed school children. The types of children studied, the classroom environment in which the research takes place, and the nature of the research were described. The development of materials and procedures was viewed as an example of applied research.

Improvement of the remedial educational process cannot rely on simply measuring behavioral change. Changes in the antecedent, consequent, and setting events that interact with behavioral change must be managed and assessed. In the three papers that follow, we will show how an effective learning environment can be arranged for the teaching of reading, arithmetic, writing, spelling, and composition skills. Although the framework for each paper deals with antecedent instructional materials, note that these are integrated with the management of setting events and behavioral consequences.

REFERENCES

Bijou, S. W. The technology of teaching young handicapped children. In S. W. Bijou & E. Ribes (Eds.), *Behavior Modification: Issues and extensions*. New York: Academic Press, in press.

Bijou, S. W. and Grimm, J. A. The education and training of the retarded child. In W. C. Becker (Ed.), *Amplifications of reinforcement principles in education*. Champaign, Illinois: Research Press, in press.

Skinner, B. F. Some relations between behavior modification and basic research. In S. W. Bijou & E. Ribes-Inesta (Eds.), *Behavior Modification: Issues and extensions*. New York: Appleton-Century-Crofts, in press.

* * *

THE INDIVIDUALIZED READING PROGRAM[1]

linda n. berner
jeffrey a. grimm

The teaching of reading involves bringing two classes of behavior under the control of printed words. The first class of behavior is the oral naming of written words, phrases, and sentences which are already in the child's speaking vocabulary. The second is reading comprehension, which involves giving a correlated response for a printed word, such as finding a picture or object, giving a verbal definition, or demonstrating an action.

This paper will describe the program used to teach oral reading and comprehension behaviors, discuss the types of monitoring data collected to evaluate the effectiveness of the program and assess the progress of each child, and present a study in which a mother learned to use the reading program at home.

DESCRIPTION OF THE PROGRAM

The tasks in the program are grouped in Table 1 according to the two classes of behavior to be taught.

TABLE 1: *COMPONENTS OF THE INDIVIDUALIZED READING PROGRAM*

Oral Reading:

1. SIGHT
    - A. Discrimination
    - B. Oral Reading/Discrimination*
2. SELF-PROMPTING
    - A. Word Forms
    - B. Blending
3. ORAL READING PHRASES*

Comprehension:

1. PICTURE IDENTIFICATION*
2. READING COMPREHENSION*
3. STORIES
4. DIRECTION TABLE

**Reading tasks given for each new word; these four tasks constitute a unit of the program.*

---

[1] We would like to thank Dr. Sidney W. Bijou for his suggestions and encouragement in the writing of this paper.

Oral Reading

The first task in the sight section of oral reading is discrimination. The ability to discriminate one word from another should be precurrent to the oral reading of words. Data on this relationship are presented in Table 2.

TABLE 2: *RELATIONSHIP BETWEEN PERFORMANCES ON THE ATTENDING PROGRAM AND THE INDIVIDUALIZED READING PROGRAM*

CHILD	ATTENDING PROGRAM DISCRIMINATION LEVEL 9, ZERO-DELAY MATCHING ACCURACY	READING-UNIT I ACCURACY
K. D.	20%	55%
B. B.	60%	71%
H. G.	40%	64%
W. J.	80%	81%
S. M.	60%	81%
K. S.	20%	81%
C. B.	40%	54%
N. B.	100%	98%
M. L. M.	100%	95%
M. K.	100%	98%

The children who performed below 90% on the discrimination program also performed below 90% on the first part of the actual reading program. Those children who did reach the 90% criterion on discrimination were successful in the beginning of the reading program.

The children in the first group, those who needed training in discriminating words, were given materials of the type shown in Figure 1. Initially, a simple match-to-sample response, in which the sample and the choices are available simultaneously, was required. This was gradually changed to a zero-delay matching task in which the sample and the choices were never available at the same time.

This discrimination task uses the same printed words which later serve as the cues, or discriminative stimuli, for the oral reading response. Reading behaviors, however, are ultimately maintained by response con-

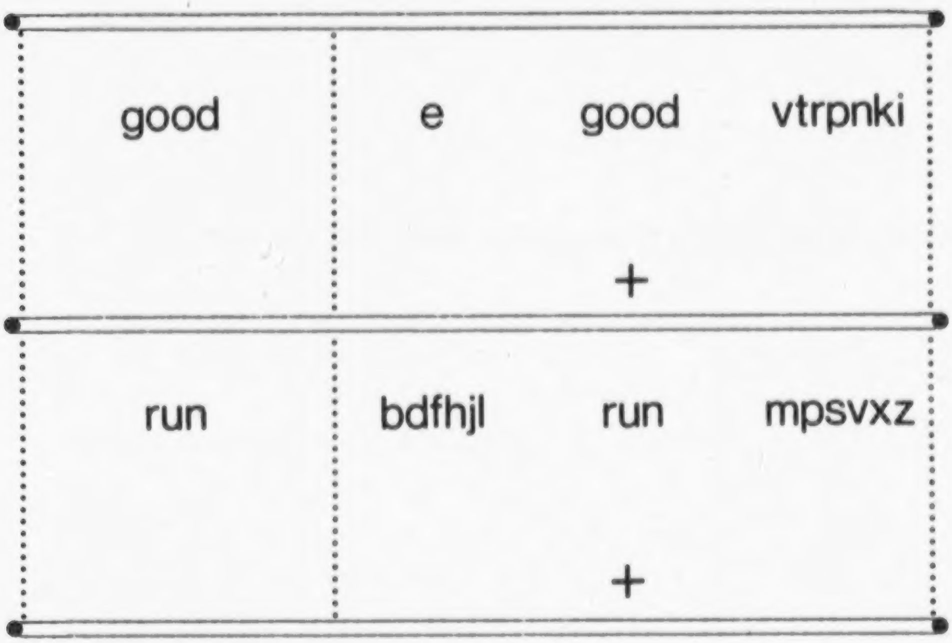


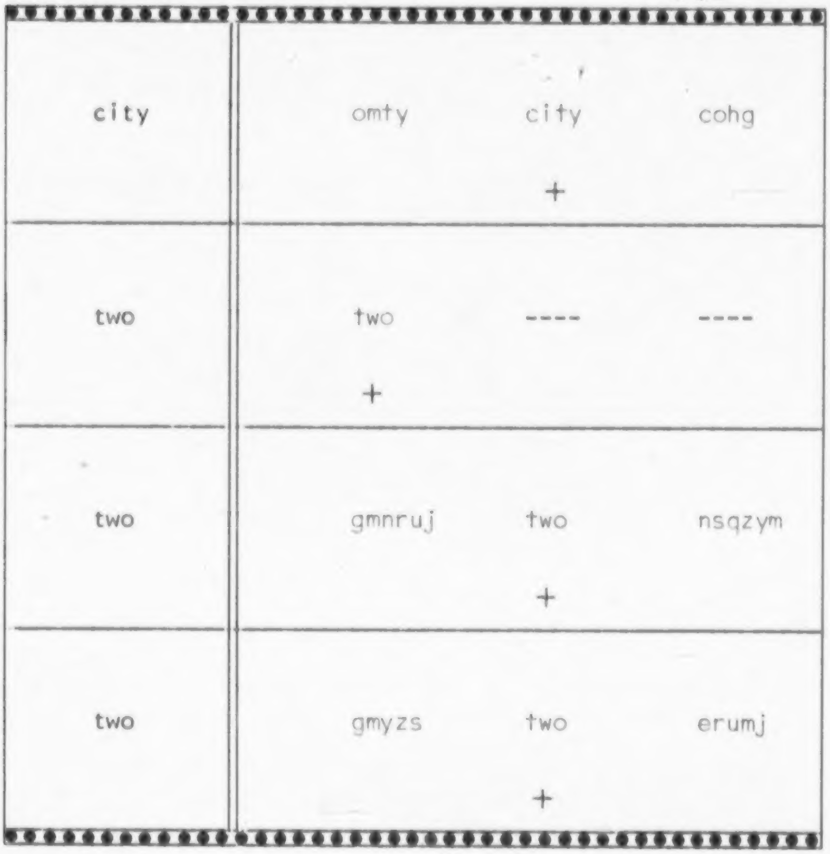
Fig. 1. Sample frames from the Discrimination Program.

tingencies such as parental approval, as well as the intrinsic reinforcers available to reading, such as the excitement of a mystery story. These contingencies are utilized as quickly as possible by rapidly providing the child with a small sight reading vocabulary. The sight reading approach allows the child to acquire quickly a reading vocabulary sufficient to demonstrate to himself and others that he is able to read.

Development of sight oral reading skills begins in the Oral Reading/Discrimination task. This sequence teaches the child the verbal equivalent of a printed word while discriminations are used for the introductory frames of the new word as well as for error correction.

On the sample page, in Figure 2, the child reads the words in the left column. If a new word is being introduced, as in the last three frames, or if any error is made in oral reading, the child is presented with a zero-dealy matching task using the words on the right.

OR/D I-4



city	omty	city +	cohg
two	two +	----	----
two	gmnruj	two +	nsqzym
two	gmyzs	two +	erumj

Fig. 2. Sample frames from Oral Reading/Discrimination.

The next component of oral reading is Oral Reading Phrases. A sample of the materials is shown in Figure 3. Here the child is presented with longer combinations of words. A phrase read in a word-by-word style is not accepted, while a phrase read in a smooth, conversational style is reinforced. In this task, capitalization and punctuation, as well as their controlling functions, are also introduced.

For a child who has not previously experienced success in reading, using his sight vocabulary skills to read meaningful phrases is a reinforcing experience.

ORP 11-9

One, two, three go.

Go up the streets and out to work.

Fig. 3. Sample frames from Oral Reading Phrases.

Self-Prompting

While the sight approach allows the child to quickly acquire a minimal vocabulary, self-prompting skills are necessary for a child to become an independent, efficient reader. The subprograms in this section (plural *-s* ending; verb endings *-s*, *-ed*, and *-ing*; and compound words) all utilize the child's present vocabulary to teach skills which can then be applied to new words. The child is also required to respond to the letter-name letter-sound relationship of initial consonants. Given a written letter, he must respond with that letter-sound and, conversely, hearing the sound of a letter in a word, he must respond with the letter-name. When a child has mastered several consonants, he is required to blend these with root words learned in the sight program. This skill is the self-prompting device which enables the child to sound out a specific type of unknown word.

Comprehension

All of the tasks discussed thus far concentrate on the development of oral reading skills. Reading, however, requires more than orally naming words. Comprehension is also necessary. The first task in this area is Picture Identification. This is the most primitive of the comprehension components, requiring only that the child respond *yes* or *no* to a question about a pictorial representation of a word (e.g., "Is this a picture of girls?").

The next task, Reading Comprehension, is shown in Figure 4. This task also uses a pictorial referent. However, the child is required to orally read the word or phrase as well as to determine whether the picture is or is not a referent for the word or phrase. He indicates his answer by circling the *yes* (smiling) *face* or *no* (frowning) *face*.

The third component of comprehension is Stories. While stories are important as a comprehension task, they also serve as reinforcers for reading behavior and as a review of known words, both in the originally taught form and in alternate forms (e.g., plural nouns or past tense verbs). The comprehension of story content is evaluated by the child's response to questions which are asked throughout each story. They are of two types: recall of specific story content or paraphrasing (e.g., "How many boys are in the store?"), and construction of answers requiring interpretation of the content (e.g., "Why do you think the boys went to the store?").

The final comprehension task is the Direction Table. Here the child is required to make an overt, physical response to written sentences. The child is asked to read aloud a direction in a booklet, and then manipulate or arrange materials in accordance with the direction given.

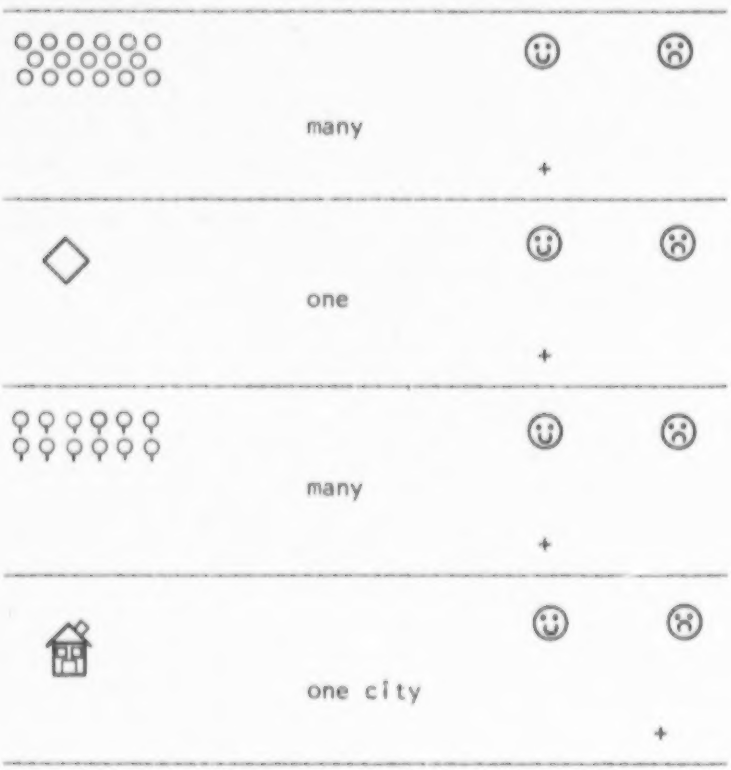


Fig. 4. Sample frames from Reading Comprehension.

For example, a written direction might be, "Put the car on the street." The child would be required to read this direction, take one of the toy cars, and place it on a model street. The manipulation of the materials is the behavior which demonstrates that the child has understood what he read. This, therefore, is the behavior which is reinforced.

Sequence of Teaching the Program

The component parts of the reading program are presented to the child in the following order. After the completion of the discrimination program, the child begins acquisition of oral reading skills through the sight program. Each new word is taught and maintained through the use of Oral Reading Discrimination and Oral Reading Phrases. The elementary comprehension tasks, Picture Identification and Reading Comprehension, are also used with each new word. These four components make up a core unit of the sight program. Self-prompting skills, which are used to expand the child's vocabulary, are introduced after the child has mastered a minimal number of sight words and when his performance on the sight program has stabilized. Stories and Direction Table are given according to the extent of the child's reading vocabulary and are programmed to become increasingly more complex.

MONITORING PROCEDURES

Each child's progress through this program is monitored daily in order to assess the effectiveness of the program and to determine the child's next assignment. The child continues to move through the program as long as he maintains the 90% mastery criterion. If his accuracy falls below 90%, a remedial unit is prepared and administered.

The following graphs are examples of this type of data. Figure 5 shows one child's accuracy on Oral Reading/Discrimination. Sixty core units of the program are represented. On both graphs in Figure 5 successive units of the reading program are plotted on the horizontal axis and percentages of correct responses are plotted on the vertical axis. The dashed line represents the 90% mastery criterion. For each point below this line, a remedial unit is administered until 90% criterion is reached on the original unit. These remedial points do not appear on the graph. The Roman numerals refer to sets of the program, each set including ten units. The upper graph shows accuracy for oral reading only. Each of the data points represents initial accuracy for a new vocabulary word combined with accuracy for a limited number of review words. Oral reading accuracy for this child has remained high throughout the school year, which indicates the efficiency of the program. The lower graph in Figure 5 shows the accuracy for the discrimination part of the task. It should be remembered that discrimination is basically an error correction procedure for oral reading. It can be seen that this child can accurately discriminate the target words and that the behavior is maintained throughout the year.

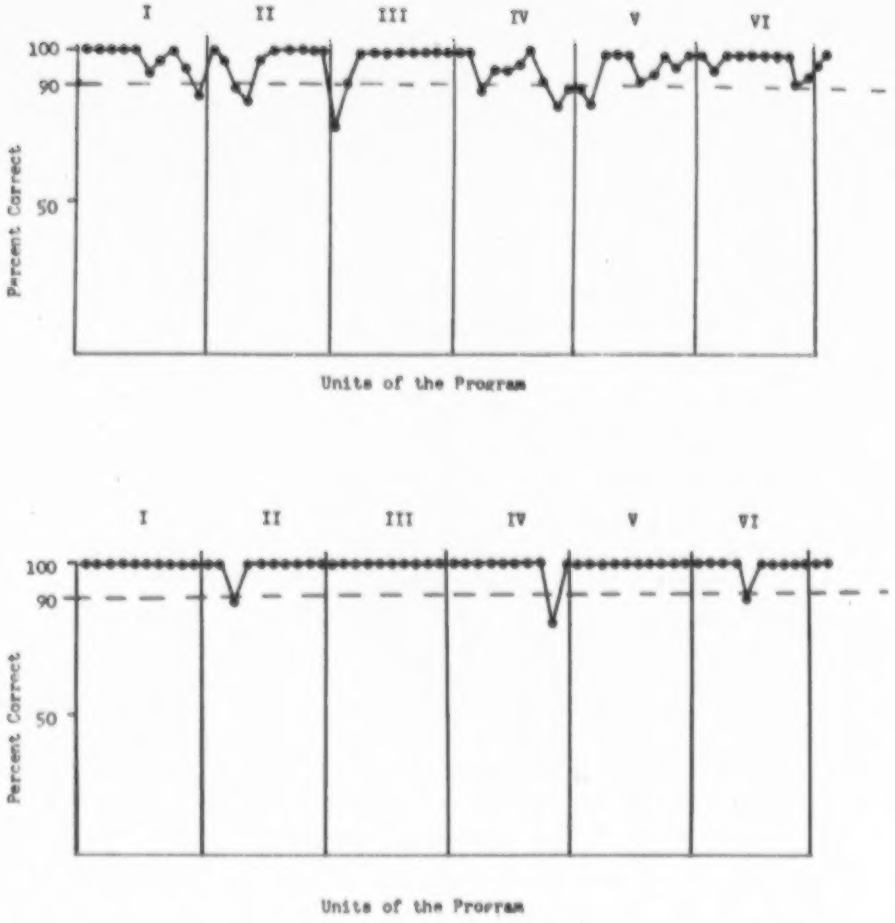


Fig. 5. Oral Reading/Discrimination.

Figure 6 shows another child's performance on Oral Reading Phrases. These graphs cover the first 34 units of the program. The top graph is oral reading accuracy. Again, no remedial units or corrections are shown, though they were administered wherever accuracy dropped below 90% criterion. This child's accuracy has remained high and stable throughout the program, with only one point falling below 90%. The lower graph shows oral reading rate. Here, the vertical axis represents the number of correct responses per minute. Decisions about reinforcement schedules are made on the basis of these data. When an increase in rate is apparent, reinforcement schedules are changed so that the child must complete more work to receive the same about of token and social reinforcement. With this child, corrections in the reinforcement schedule were made frequently in Units III and IV since oral reading rate was increasing.

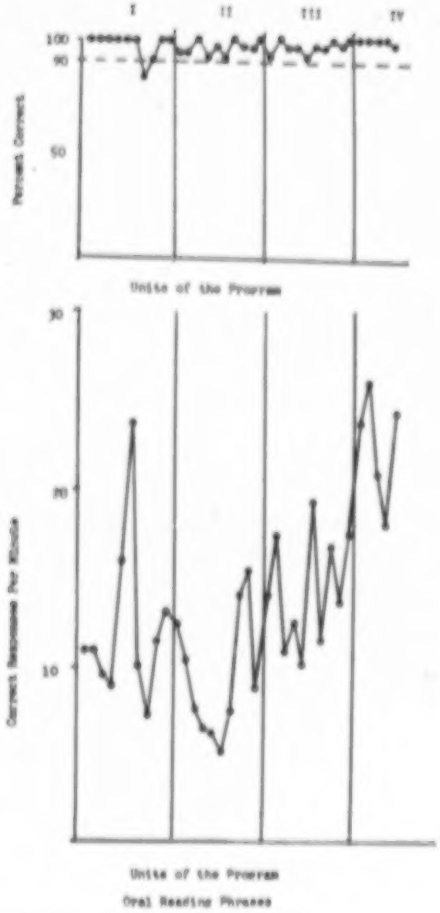


Fig. 6. Oral Reading Phrases.

The data in Figure 7 show a third child's performance on the Reading Comprehension task. The graphs begin after the child has completed 40 units of the program and continue through 100 units of the program. Both graphs show accuracy, the upper one for oral reading, the lower one for comprehension. These two measures are separated since it is important to know how the child is progressing in both areas. For example, while this child does very well in oral reading, she has difficulty with comprehension. For most errors in comprehension, remediation is performed immediately and the frame is repeated. Only the child's first response, however, is recorded. More formal remedial units are given only when the accuracy remains consistently low for the same word or phrase. The points labeled ☐ on the graphs represent core units that were not given to the child because she could read and define the word taught in those units on the pretest. The

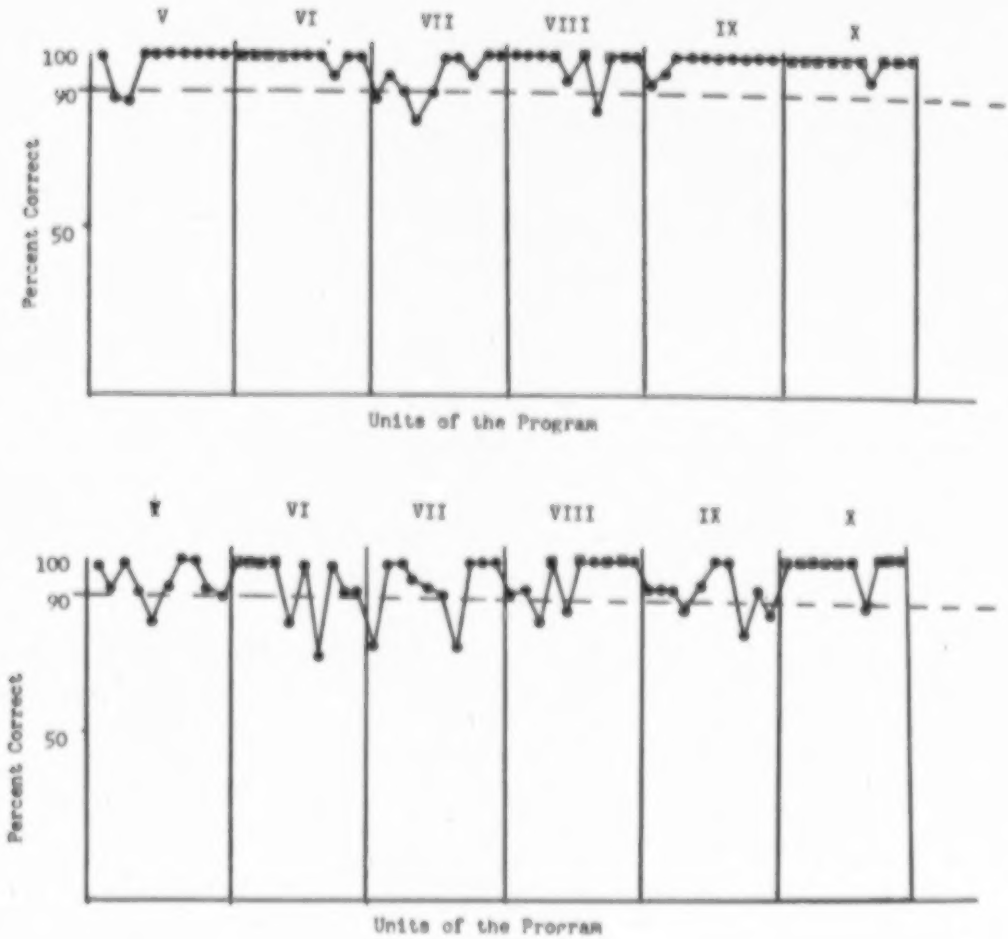


Fig. 7. Reading Comprehension.

frequency of these points increases as the child progresses through the program because she is simultaneously acquiring self-prompting skills. Reliability on the collection of this monitoring data is taken periodically by independent observers. It is established and maintained at 90% throughout the year.

Remedial Units

The data presented thus far are used as a basis for making decisions about the child's academic progress and about reinforcement scheduling. A more detailed example of the monitoring and remediation process is shown in Figure 8.

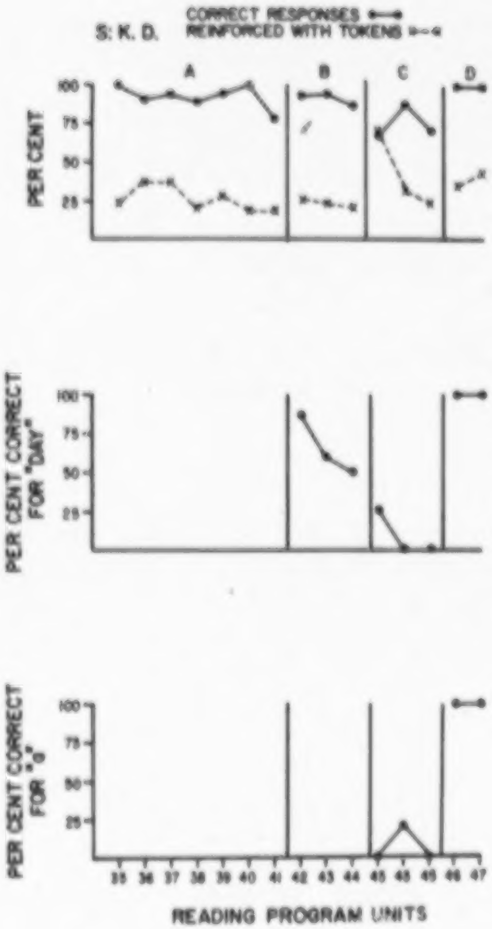


Fig. 8. Combined Oral Reading Phrases and Oral Reading/Discrimination accuracy before and after a remedial unit and the percent of correct responses reinforced with tokens for all words (top graph). The combined Oral Reading Phrases and Oral Reading Discrimination accuracy before and after a remedial unit for the words "a" and "day" (middle and bottom graphs respectively).

Units of the reading program (numbered on the lower graph) are represented on the horizontal axes. The vertical axes show percentages. The solid curves indicate percent of correct responses, and the dashed curve indicates frequency of token reinforcement. The A portion of the top graph shows that the child's performance on oral reading and comprehension tasks was generally accurate and stable after 41 units of the program. The dashed curve shows that token reinforcement was about 25%. Unit 42 introduced the word *day*. The middle graph shows accuracy for this word only. It can be seen in the *B* portion of the top and middle graphs that overall accuracy remained stable and accuracy for the word *day* decreased across Units 42 through 44. Assuming that the low rate of token reinforcement (seen in the broken line) might have been related to the decrease in accuracy, the frequency of token reinforcement was increased with the introduction of Unit 45 (*C* portion of the graphs). Unit 45 introduced the new word *a*, whose accuracy is shown in the lower graph. Even with increased reinforcement, the accuracy for *day* decreased and the word *a* was not acquired. Repetition of Unit 45 did not result in mastery of either word. However, during this repetition, the child usually responded with the verbal *day* to the written word *a* and vice versa. Therefore, a remedial unit was constructed that emphasized discrimination between the written form of *a* and *day* and utilized a much-condensed version of the normal transition from the reading discrimination task to the oral reading phrases task. Following the successful administration of this unit, the child went on in the reading program. Portion *D* of all graphs shows that the child's accuracy rose immediately to 100% for the words *day* and *a* as well as for the new words in Units 46 and 47.

Posttest Evaluation

It has been mentioned that a systematic review of previously acquired words is used to maintain oral reading behavior. The usefulness of this process can be seen in the posttest data shown in Table 3 for all

TABLE 3: *READING POSTTEST ERRORS AS A FUNCTION OF NUMBER OF REVIEW SESSIONS*

UNIT NUMBER	NO. OF REVIEW SESSIONS	PERCENT ERRORS ON POSTTESTS
1	3	1.63
2	3	1.63
3	3	2.44
4	3	2.44
5	3	.81
6	3	.81
7	3	1.63
8	2	4.88
9	1	6.50
10	0	12.20

children on the reading program. These data are based on 123 separate posttests. A posttest is given after every set of ten units of the program and covers the ten words included in that set. The word taught in a given unit appears frequently in all of the component tasks of the next three units. Subsequently it is systematically reviewed in the Reading Phrases and Reading Comprehension tasks only. Therefore, at the time of the posttest, the words taught in Units 8, 9, and 10 have had the least amount of review. The table shows that for those words which have had at least three reviews, the average percent of errors is 1.63. When review sessions drop to zero, the percent of errors increases to 12.20. Therefore, it appears that the review is functional as a maintenance device. The data also show that posttest errors in general are low in frequency, ranging from less than one percent to about twelve percent as review is decreased. This is an indication of the efficiency of the program.

PARENT TEACHING

The program's use is not limited to the classroom and its staff. Figure 9 shows a hyperactive child's progress on the program. The open circles represent the cumulative number of correct responses completed and the closed circles represent the number of incorrect responses. Each data point represents reading output summarized over three reading days. Days in which no reading occurred (i.e., weekends and holidays) are not represented. During the period marked *A* the child read exclusively during a 25-minute tutored reading period at school. The rates shown are typical of the other children on the program.

At point *B*, the child's mother, who had a past history of ineffective interaction with her son, began to tutor him at school. The child's performance was unaffected; both correct and incorrect responses remained stable.

During *C*, the child worked with his mother at home as well as with his regular reading tutor at school. The rate of correct responses doubled. During the 30 days of *C*, the child made 4000 correct responses. During the 66 days of *A*, the child made 4500 correct responses. Rate of incorrect responses remained relatively constant.

AA represents a second period during which the child read only at school with his tutor. His performance was comparable to that in condition *A*. In *CC*, his mother again tutored him at home and performance increased as it had in period *C*. These data suggest that a mother can be trained to use this complex reading program at home with results that significantly augment the teacher's efforts in the classroom.

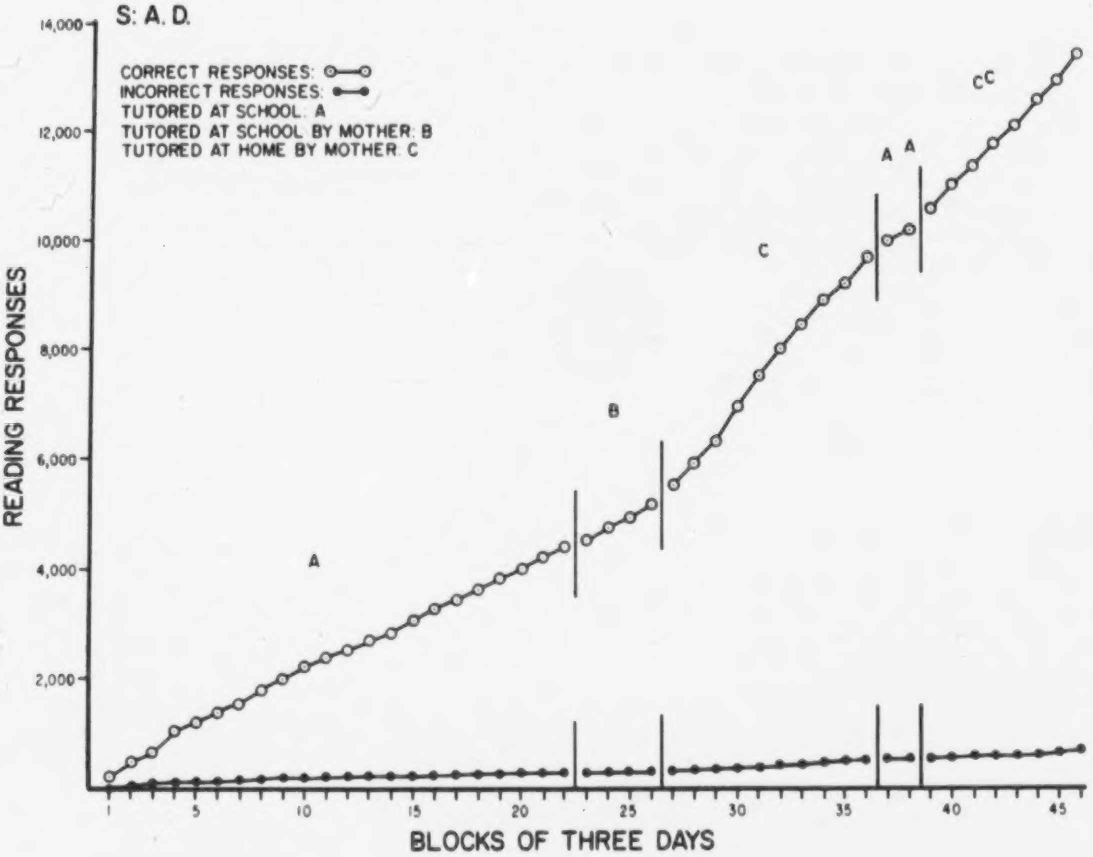


Fig. 9. Cumulative record of the progress of A. D. in completing the Individualized Reading Program.

SUMMARY

Research on reading has been concerned with both oral reading and comprehension behaviors. The resulting Individualized Reading Program utilizes a combined sight and phonetic approach in teaching oral reading. Sight reading tasks enable the young handicapped child to experience success in reading and, hence, access to terminal reinforcers as quickly as possible. However, it is realized that the child must eventually acquire self-prompting skills if he is to become an efficient, independent reader. Reading comprehension is taught in four types of tasks. The most primitive of these requires the child to identify a picture referent for a word and the most complex requires him to follow written directions without assistance from the teacher.

Evaluation of the program's effectiveness and each child's progress is accomplished by monitoring performance on each reading task. Data for this purpose are oral reading and comprehension accuracies, oral reading rates and scores on pre/posttests. This information serves as a basis for curriculum planning for each child (e.g., remedial units and reinforcement schedules). The program has also been successfully used by parents outside the classroom situation.

APPLICATION OF BEHAVIOR PRINCIPLES TO THE TEACHING OF WRITING, SPELLING & COMPOSITION[1]

ely rayek
elizabeth nesselroad

This research is concerned with the development of instructional materials and procedures for teaching writing, spelling, and composition to young handicapped children. Although each of the three programs has distinctive objectives and incorporate different procedures, they are functionally interdependent. The program progression is designed so that writing is maintained by spelling and spelling is maintained by composition.

This paper describes the research on the writing, spelling, and composition programs in terms of the analysis of the task to be taught and the procedures explored in arranging conditions to expedite their learning. Samples of data will also be presented.

WRITING

The writing program has two terminal behavioral objectives:

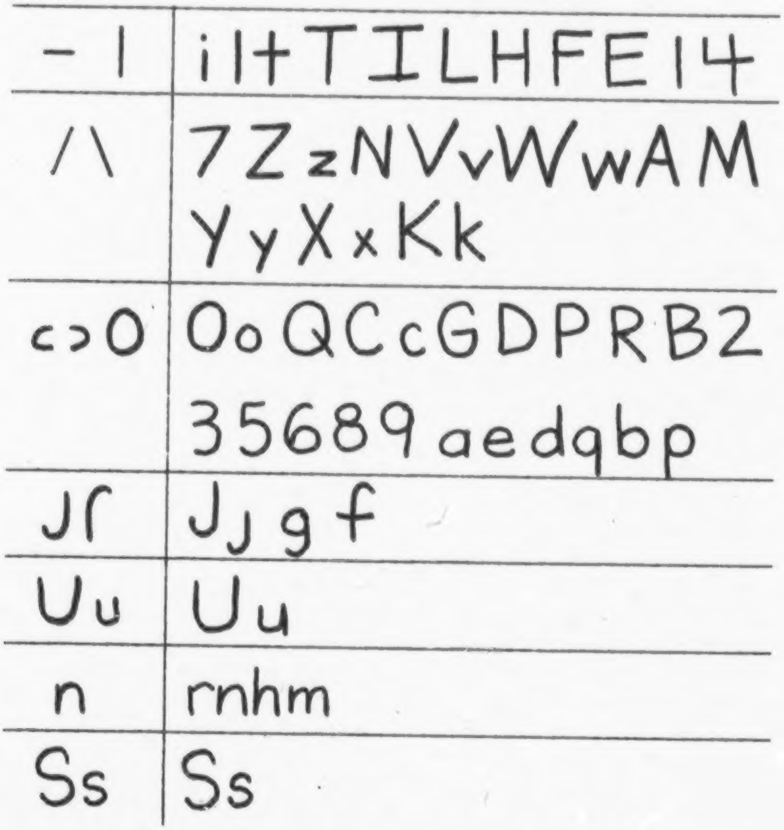
(1) write from dictation all of the upper and lower case letters of the manuscript alphabet and all numbers from 0 to 9; and

(2) write and blend from dictation all of the upper and lower case letters of the cursive alphabet.

In constructing a program to teach writing to young handicapped children, several points are considered. First, writing is taught by shaping procedures. It is the gradual shaping of a motor skill to the point where the response approximates a specified model. Second, letters are made up of stroke elements. A task analysis suggests that the elements requiring the least complex motor responses are the horizontal and vertical lines. Slanted lines and curved lines are more complex. For maximum program efficiency, letters containing common elements are grouped into families and taught together. For instance, all letters using straight lines are taught before letters containing slanted lines. According to the task analysis, letters within a given family are ordered according to their difficulty and assigned in that order. Figure 1 shows the stroke elements and the letters and numbers which include them. Third, copying a letter from a visual model precedes writing a letter from dictation. This procedure is used to assure that the child is equipped to make the appropriate strokes prior to writing from dictation. Fourth, writing is functional when it is used in practical ways such as writing one's name, words, or stories. Once these meaningful uses of writing are possible for the child, the social interactions which ensue take over to maintain the behavior.

---

[1] We wish to express our appreciation to Dr. Sidney W. Bijou for his guidance in this research and his editorial assistance with the manuscript.



LETTER FAMILIES

ELEMENTS	LETTERS AND NUMBERS
- |	il+TILHFE14
/\	7ZzNVvWwAM YyXxKk
c>O	OoQCcGDPRB2 35689 aedqbp
Jſ	Jjgf
Uu	Uu
n	rnhm
Ss	Ss

Fig. 1. Elements and letter families.

Manuscript Writing

There are two prerequisites to the terminal manuscript objectives:

① pencil holding; and

② discrimination of well-formed and poorly-formed letters.

Children who at the beginning of the school year do not hold the pencil correctly are given instruction by means of a modified modeling procedure. When this procedure is ineffective, a colored dot is placed on the pencil and on the corresponding part of the child's hand at every point where the child's hand should come in contact with the pencil. The child then matches the colored dots to achieve the correct position for holding the pencil.[2]

The ability to respond differentially to poorly-formed and well-formed manuscript letters is a second important prerequisite which enables the child not only to correct and to reinforce his own written responses but to work independently. A subprogram was designed to teach the child this skill by requiring him to discriminate between poorly-formed and well-formed letters. In the initial frames the child is required to match well-formed letters. Example frames are shown in Figure 2.

The core writing program requires four different classes of responses:

(1) copy the letter from a visual model;
(2) write the letter from dictation;
(3) recognize the letter; and
(4) name the letter.

Letter recognition is a prerequisite for letter naming.

Procedurally, the program for copying letters from visual models utilizes (1) the giving and fading of prompts, and (2) the reinforcement of progressive approximations to the desired end-product. Prompts used are of two types: *visual* and *verbal*. Visual prompts consist of line and dot cues which the child is required to trace. Initially, the child must trace a model of the letter. Then, he is required to trace models of the letter which are in dot format (e.g., for T). Fewer dots are provided for subsequent responses until, terminally, there are no dots provided. Verbal prompts are instructions which pre-

---

[2] More detailed information is available on instruction in pencil holding in the Final Report to USOE, August 1971, entitled "Application of Behavioral Principles to the Remedial Instruction of Retarded and Emotionally Disturbed Young Children."

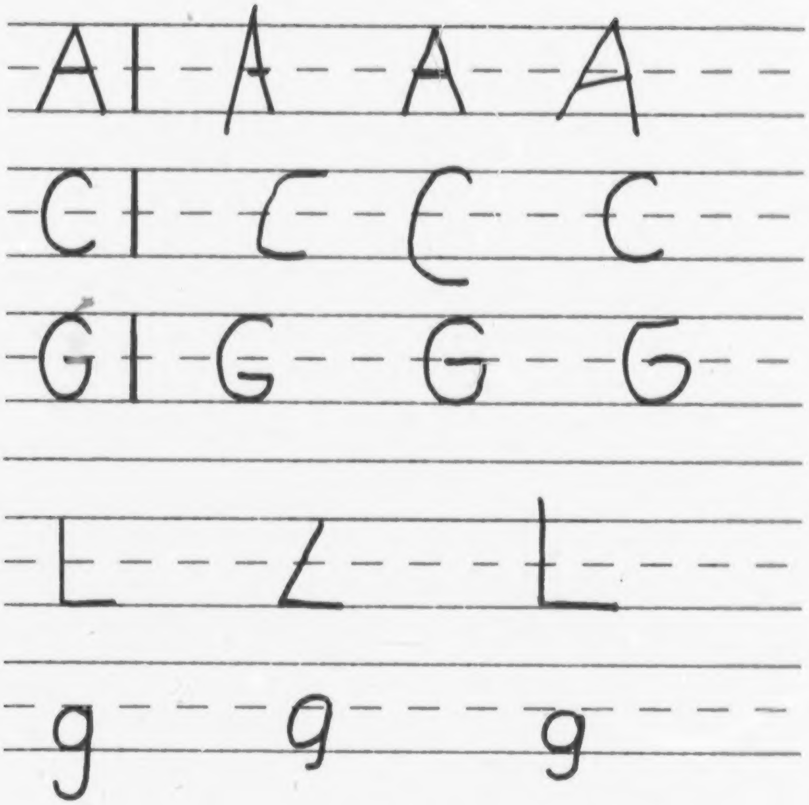


Fig. 2. Sample of initial frames used in the program to teach the discrimination of well-formed letters.

cede a response and might be, "Make a straight slanted line" or "Make the circle smooth."

After an incorrect copying response, the teacher makes changes on the child's response to make it consistent with the model (e.g., straightening crooked lines or drawing the letter between the lines). A new letter is begun, sometimes with dot prompts, sometimes without depending on the child's previous responses for that letter. When a child has demonstrated that he can make a well-formed letter but he is not consistent in writing the correct response, an appropriate verbal prompt is given prior to the next response. However, since verbal error correction procedures have reinforced incorrect responding for some children, verbal interaction or verbal prompts are kept to a minimum. When they are used, it is always in the context of the next response and not of the one on which an error was made. After the child copies the letter at least three times in recognizable form, visual models are eliminated and he is required to write that letter from dictation.

Procedurally, the dictation task consists of the teacher covering or removing all visual models and saying to the child, for example, "Make an *i*." When the child makes an error (makes a different letter than the one he is instructed to make) the teacher provides a model and instructs the child to copy it. If the child writes the correct letter from dictation but the form is incorrect, the same error correction procedure is used as on the copying responses. After the child makes a recognizable form of the letter on three consecutive attempts, that letter is mixed with other previously mastered letters and he is required to write them from dictation. This is done to assure that the child's response is under the control of the name of the letter and is not a result of repetition.

Scoring of letters is according to two sets of criteria. First, all letters written (both from a model and from dictation) are scored according to letter form. Letter form is evaluated according to specific criteria. The letter must be within and just touching the guidelines; horizontal, vertical and slanted lines must be straight; curved lines must be smooth. Samples of acceptable forms are given in the upper part of Figure 3. Criterion for mastery of letter form is two consecutive sessions with 90% accuracy. Scoring reliability is obtained by an independent observer and is maintained about 90% throughout the program. Figure 4 illustrates the number of responses to criterion for a child on writing the sequence of letters. For the first letter taught, the lower case *i*, the child required more than 300 responses to reach criterion. For the number four, which is made up of three strokes, the child required less than 40 responses to reach criterion. It is interesting to note that the number of responses to criterion for letters with slanted strokes remained low in comparison with the number of responses to criterion for the initial letters of the first family.

The second set of scoring criteria is for dictation responses which are scored as either correct or incorrect in terms of whether a letter is recognizable as the assigned one. An example of acceptable (or recognizable) letters for the dictation response is given in the lower part of Figure 3. Although accuracy for individual letters is computed, all letters written from dictation in a single period are combined to derive an overall accuracy score. Typically, children in our program maintain an accuracy nearly always above 90% even though new letters are being introduced daily.

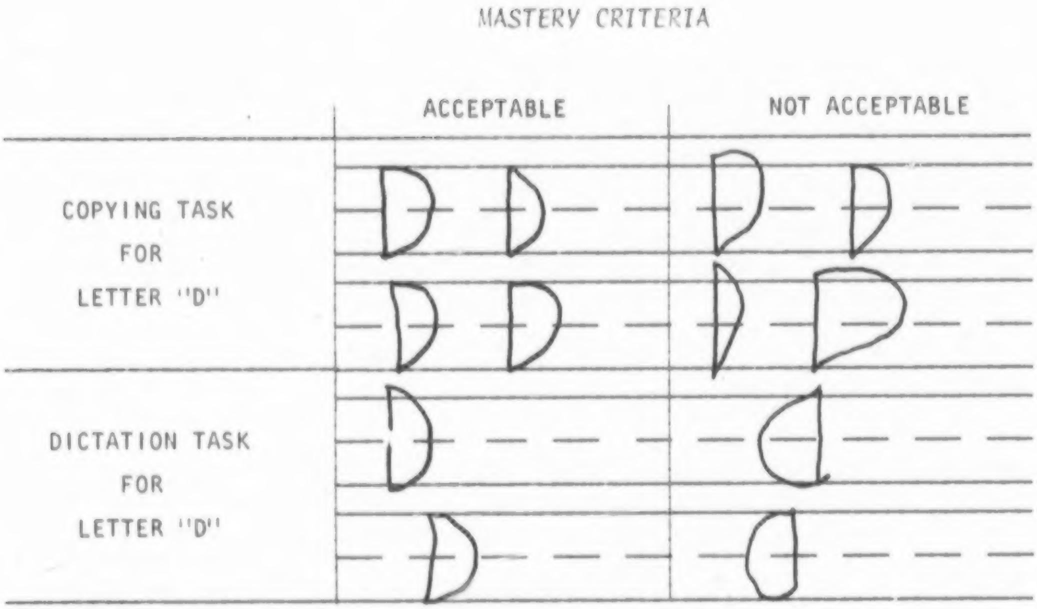

MASTERY CRITERIA
ACCEPTABLE
NOT ACCEPTABLE
COPYING TASK FOR LETTER "D"
DICTATION TASK FOR LETTER "D"

Fig. 3. Examples of both acceptable and unacceptable letters for the dictation and copying responses.

Letter naming is taught simultaneously with copying a letter for two reasons. First, since mastery of letter form requires many copying responses, there is good opportunity for the child to recognize and name the letter. Second, writing from dictation, taught after the child's initial letter naming responses, appears to be facilitated by the letter naming skill.

At the beginning the child is required to recognize the letter (e.g., "Is this an *A*?" or "Show me an *A*."). Later he is required to name the letter (e.g., "What is the name of this letter?"). Within each class period, the child is required to name or recognize more than one letter. As with the dictation task, children maintain high accuracy (above 90%) for letter naming. When the child makes an error

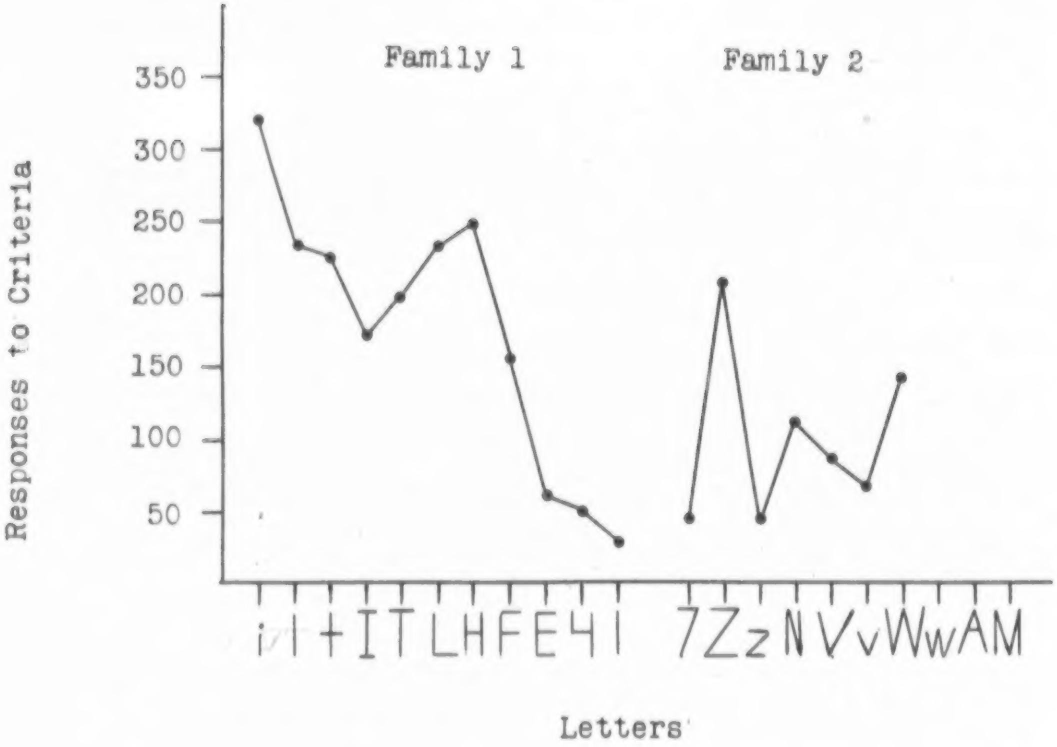


Fig. 4. Sample data of a child's performance in writing letters.

on naming, recognition, or dictation tasks, he is corrected, prompted, and asked to repeat the response.

Many opportunities to respond to copying, dictation, and naming tasks are provided in each writing period. For instance, the teacher asks the child to copy an *A*. She reinforces his response and then says, "Is this a big *A*?" (the recognition task). Another copying response may be followed by an opportunity to name the letter. After the child makes a recognizable form of the letter, the teacher might eliminate all visual prompts and ask him to write the letter from dictation.

Cursive Writing

A cursive writing program is available for children who complete the manuscript program. Only those children whose responses in the manuscript program are both accurate and rapid are assigned the new program. In addition, only children for whom the new skill would be maintained upon their return to public schools (children going into second or third grades) are introduced to the cursive program.

As well as being able to name, copy, and write from dictation the letters of the alphabet, the child is taught to write a cursive letter from a manuscript model. Procedurally, the child is presented with both a cursive and manuscript model. (Figure 5-A) He is in-

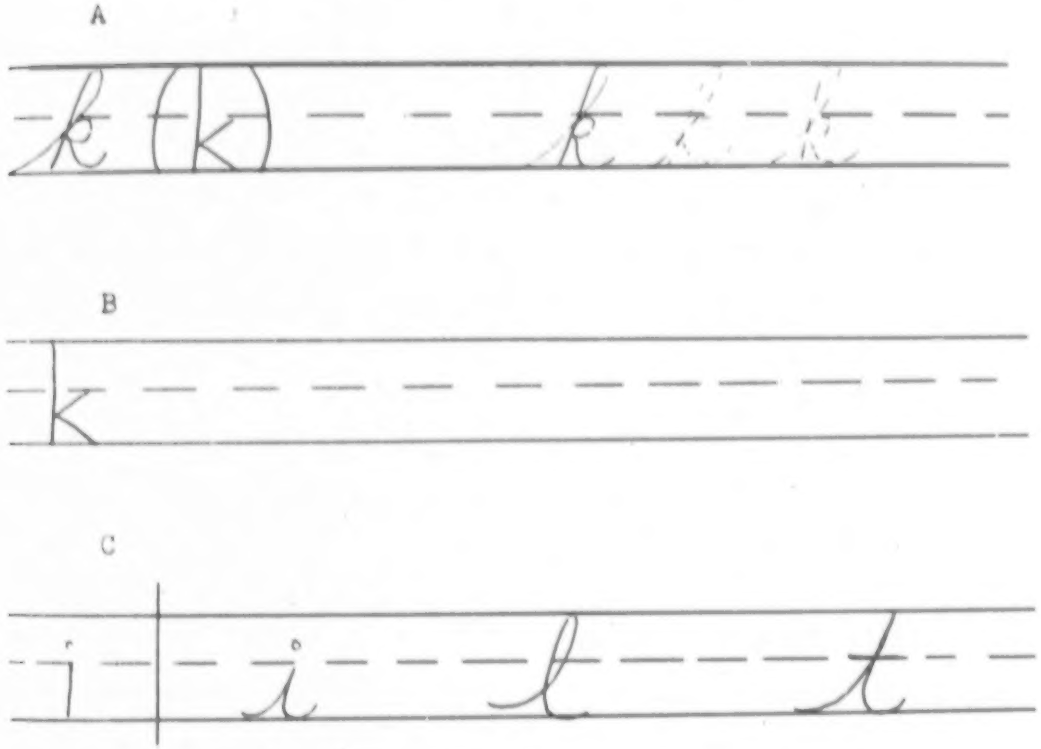


Fig. 5. *A* and *B*. Sample frames of the program used to teach the equivalence of manuscript and cursive letters.

C. Sample frame of the remedial program used to teach the equivalence of manuscript and cursive letters.

structed to name the letter and make several copying responses. When the child can make a recognizable form of the cursive letter, he is presented with just the manuscript letter model and instructed to, "Write this letter in cursive style." (Figure 5-B) On letters for which this procedure is ineffective, the child is given remedial frames like those shown in Figure 5-C. He is instructed to read the letter on the left, cover it, and circle and name the letter with the same name.

The cursive program concludes with a procedure for letter blending. The program proceeds from practice in copying groups of two letters to copying words. Terminally, the child is able to write such letter combinations from dictation.

SPELLING

Spelling is analyzed as the oral or written production of the letters of words in a given sequence under the control of spoken stimuli. The skills acquired in the writing program, those of naming and writing the letters of the alphabet, are prerequisites for the spelling. The terminal objective of the spelling program is that the child spell - in both oral and written form - all of the words in his reading repertoire.

The sequence in spelling is shown in Table 1. First the child reads the word and the letters of the word from a study card. Then in a delayed matching sequence, he looks at the word, turns the card over so that the model is hidden, and immediately writes the word. After he has written the word, he exposes the model and checks his spelling. In a single period, these first three steps in the sequence are completed for all words assigned (usually five words in a 20-minute period). Then the teacher removes all printed material and requires the child to spell the word orally. The teacher says, for example, "Spell *city*." After the child orally spells all assigned words, he is required to write the words from a spoken model (e.g., "Write *city*."). After the child writes each word, the teacher checks the child's work.

The child goes through different error correction procedures depending on the errors he makes. The column on the right side of Table 1 shows the error correction procedures. If the child incorrectly reads the word, the teacher corrects him and prompts him to read it correctly. If the child incorrectly reads any letter of the word, the teacher stops him immediately, corrects him, and requires him to give the spelling response chain from the beginning. The child corrects his own spelling in step 3. If he is incorrect, he has to repeat the delayed matching sequence, writing the word again in the absence of a printed model. If the child makes an error while spelling the word orally, the teacher stops him and gives the correct sequence. Then the child is required to spell the word from the beginning once

SPELLING PROGRAM

The child:

NO ERROR SEQUENCE		ERROR CORRECTION PROCEDURE
① Reads the word from a study card.	*city*	① Teacher corrects the child and prompts him to repeat it.
② Reads the letters of the word from a study card.	*c-i-t-y*	② Teacher stops him, corrects him, and requires him to start again.
③ Turns card over and writes the word.		③ Child looks at the card, checks his spelling, and, if incorrect, writes it again.
④ Spells word orally from spoken model.	*"spell city"*	④ Teacher stops him, corrects him, and he starts again.
⑤ Writes word from a spoken model.	*"city"*	⑤ Teacher writes the word and asks the child to copy it.

TABLE 1: Left Column: Sequence for acquiring spelling.
Right Column: Error correction procedures at each step of the sequence for acquiring spelling.

more. In the final step, if the child writes the word incorrectly from a spoken model, the teacher provides him with a handwritten model of the word and requires him to copy it. This lower level task assures that the child is re-exposed and reacts to the correct response after he has made an error. The behaviors required in steps 4 and 5 are the terminal objectives of the program. The criteria for mastery of any word is that the child spell it correctly in both oral and written form from a spoken stimulus in two consecutive periods. The child's progress (and as a consequence the effectiveness of the program procedures) is monitored by recording the number of words mastered daily.

Figure 6 shows the cumulative records of two children in this program. This figure shows that as the children move through the program their progress becomes rapid and stable. It will be noted that the child's performance represented in the top graph shows that he mastered about two words per day. The lower graph shows the progress of a child who mastered about one word per day. Each child had two 20-minute periods of spelling each day.

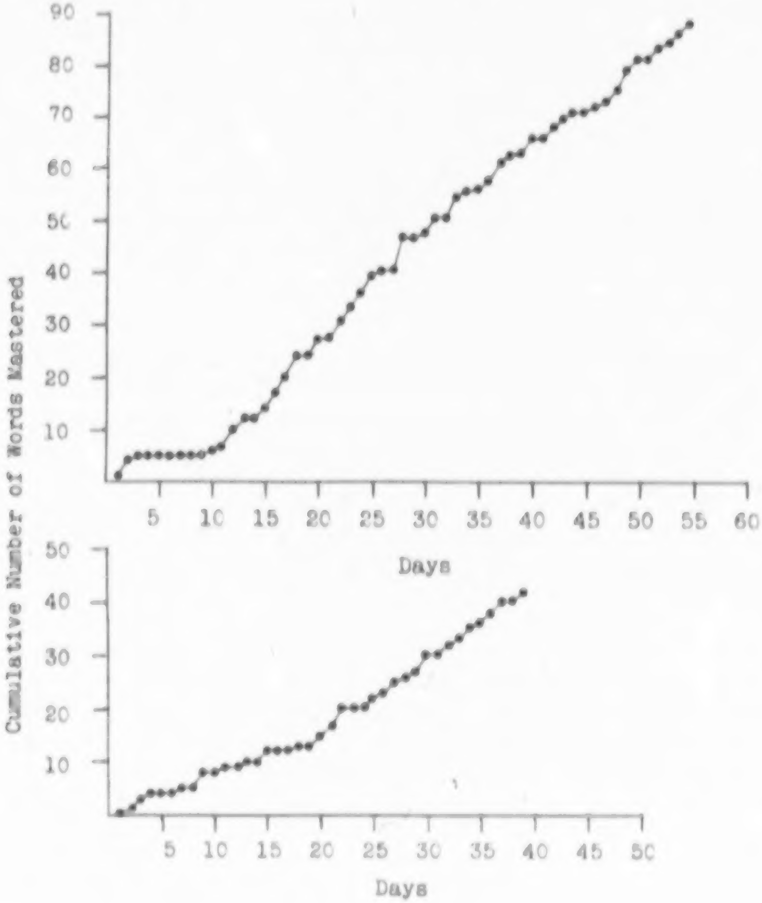


Fig. 6. Sample performance of two children on the spelling program. (Horizontal axes represent days in the program. Vertical axes represent number of words mastered.)

COMPOSITION

Composition is viewed here as an extension of writing and spelling where the production of phrases and sentences is under the control of pictures, questions, directions or the child's own verbal behavior.

Composition, in its terminal form, is a constructed or creative behavior. Specifically, it must fulfill two requirements. First, it must conform to the conventional rules of grammar. Second, the behavior must be functional: it must evoke a characteristic response from the reader.

The terminal behavioral objective for the composition program is that the child produces meaningful phrases and sentences in response to general directions. There are three prerequisites to this program. First, the child must demonstrate minimal skill in sentence construction. This is an extension of the spelling and language programs and is taught simultaneously with composition. Second, he must have the necessary reading and spelling skills to embark upon the composition program. The number of words a child can read and spell determines the complexity of questions and answers used in this program for that child. Third, the child must be proficient in writing groups of words from dictation. This skill is an extension of the requirement of the spelling program and is mastered by the child prior to his beginning the composition program.

In the initial phase of the composition program the child is required to respond to both a picture and a question. In the presence of the teacher, the child reads the question, looks at the picture, and responds orally to the question. He is then instructed to write the answer to the question. It need not be a complete sentence. The main concern is that the child's response "make sense." After the child becomes proficient in this skill, the same material is presented again with the requirement that the child's response be in an acceptable sentence form.

In the third step, material of a similar type is presented but the teacher is no longer present to respond to the child's reading of or oral answer to the question. The only behavior to which the teacher responds is the written product. Although errors in spelling and grammar are corrected, the child is encouraged to use new words even if he misspells them. For example, a response containing spelling errors would be reinforced and then spelling errors would be corrected. A grammatically correct answer with no spelling errors which is not an appropriate answer to the question would not be reinforced.

In the fourth phase of the program, the pictorial stimulus is eliminated. The child is asked a specific question such as, "How do you go to school?" The response is judged in the same manner as in the previous unit.

In the fifth phase of the program, the specific question is replaced by general directions. Initially, pictures are again presented and are then faded. An example of the initial general directions might be, "Write one sentence about the picture." An example of general directions in the last phase of the program might be, "Write one sentence about your family." More advanced units require the child to write an increasing number of sentences about the picture.

Notice that this program begins with multiple cues for the child's response which are gradually reduced. In the initial phase of the program, there are three observable cues: the picture, the question, and the child's oral response to the question. In the last phase of the program, there is one cue: the directions.

On the composition pretest, one child in our program was unable to give even a verbal response to the question-picture frames. After completing the program just described, the child responded consistently with stories of four or more sentences on assigned topics.

SUMMARY

This paper describes the research on the writing, spelling and composition programs in terms of the analysis of the task to be taught and the procedures explored in arranging conditions to expedite their learning.

The writing program which consists of shaping procedures has two terminal objectives:

(1) write from dictation all of the upper and lower case letters of the manuscript alphabet and all numbers from 0 to 9; and

(2) write and blend from dictation all of the upper and lower case letters of the cursive alphabet.

The core writing program requires four different classes of responses:

(1) copy the letter from a visual model;
(2) write the letter from dictation;
(3) recognize the letter; and
(4) name the letter.

The cursive program requires, in addition, writing cursive letters from their manuscript models. The cursive program concludes with a procedure for letter blending. Spelling is analyzed as the oral or written production of the letters of words in a given sequence under the control of spoken stimuli. The terminal behavioral objective of the spelling program is the child spelling in oral and written form all of the words in his reading repertoire. The core of the spelling program requires the child to (1) read a word, (2) read the letters of the word, (3) write the word without the written model present in a delayed matching sequence, (4) spell the word orally, and (5) write the word from a spoken model.

Composition is viewed as an extension of writing and spelling where the production of phrases and sentences is under the control of pictures, questions, directions or the child's own verbal behavior. The terminal behavioral objective of the composition program is that the child produce meaningful phrases and sentences in response to general directions. The core of the composition program requires the child to:

1. read a question, attend to a picture and orally respond to the question in the presence of the teacher. He is then instructed to write the answer. *It need not be a complete sentence.*
2. read a question, attend to a picture and orally respond to the question in the presence of the teacher. He is then instructed to write the answer in sentence form.
3. read a question, attend to a picture and write *the answer in sentence form.* From this phase on the teacher only responds to the child's written answer.
4. respond to a specific question.
5. respond to general directions.

Program progression is designed so that writing is maintained by spelling and spelling is maintained by composition.

◆ ◆ ◆ ◆

THE RECIPROCAL MODIFICATION OF ARITHMETIC BEHAVIOR AND PROGRAM DEVELOPMENT[1]

joseph a. parsons

Arithmetic behavior may be analyzed as operant behavior (Skinner, 1957). Table 1 illustrates the three-term interactional model (contingency) as it applies to arithmetic. Across the top are shown the three components of the contingency: *antecedent events, arithmetic behavior,* and *consequent events.* Antecedent events refer to the cue or discriminative stimuli to which the child responds. Specifically, antecedent events include countable objects, written or spoken numbers, and written or spoken expressions of mathematical operation. Arithmetic behavior, shown as the middle term, is comprised of counting or derivational responses. We might say someome is performing arithmetic if he is manipulating sets of objects in reference to quan-

TABLE I

THE THREE-TERM CONTINGENCY IN ARITHMETIC

Antecedent Events ⟶ (S^D)	*Arithmetic Behaviors* ⟶ (R)	*Consequent Events* (S^R)
① Countable objects	① Manipulation of sets	① Social reinforcement
② Written & spoken numbers	② Saying numbers and signs	② "Intrinsic" reinforcement
③ Written & spoken expressions of operation	③ Writing numbers and signs	③ Access to reinforcers or escape or avoidance of aversive stimuli
$\begin{array}{r}2\\ \underline{+2}\end{array}$	*"Four"*	*"That's right!"*

---

[1] The author wishes to thank Drs. Sidney W. Bijou and Jeffrey A. Grimm for offering invaluable comments on the research and writing of the manuscript.

tity. Saying and writing numbers and operations are also arithmetic behavior. The complexity of arithmetic is due to the length of behavior sequences (response chains) and not because the response elements themselves are difficult to perform. One need only emit these relatively simple responses under the appropriate conditions. Complex problems require longer chains and more varied response elements. Designing instruction for arithmetic, then, involves stimulus control procedures more than response shaping. Therefore, the interaction between antecedent events and the behavior will be of major importance. Consequences which strengthen and maintain arithmetic behavior are reinforcing events. For arithmetic, as for other classes of verbal behavior, the reinforcers are primarily social in nature. Most commonly this includes praise from parents and teachers (e.g., "Good work," "That's correct," etc.). The child may also be reinforced by manipulating the materials themselves as in playing with an abacus, blocks, or playing cards; and a great deal of arithmetic is maintained by the access to positive reinforcers, or escape or avoidance of aversive stimuli (e.g., exchange of money for goods and services, telling time to make a rendevous, driving below the speed limit).

PROGRAM ANALYSIS OF ARITHMETIC

This paper discusses the development of a program to teach arithmetic to young developmentally retarded children. It is a beginning program, requiring only a verbal imitative repertoire. Children may enter the program at higher levels depending on pretest assessment of their entering repertoires. The terminal behavior of the present program is the solution of story problems, and is best described with an example item. Miss Castle has fourteen girls and twelve boys in her class. Only five of the children can paint at one time. How many children must wait to paint? Solving this problem requires many skills. Obviously the child needs reading comprehension and writing skills as well as a complex computation repertoire. Note that both addition and subtraction operations are required to solve the problem.

In terms of the child's total learning environment, the most important reason for the selection of story problems as the terminal objective is that story problems, more so than straight computation, approximate making arithmetic functional for everyday living as well as the solution of other academic problems (e.g., physics, chemistry, etc). Just as reading, writing, and spelling are useful for and maintained by solving story problems, so are story problems prerequisite or precurrent to solving problems likely to be encountered in non-school settings (e.g., buying groceries).

Our children's arithmetic repertoires are assessed on enrollment to the Laboratory school. The eight children in the class could all verbally imitate the numbers 1-20, but could not count these numbers reliably or even name the numbers. The task, then, was to design a sequence of instruction which would lead the children from their limited entering repertoires - imitation of numbers - to the terminal objective, solving complex story problems. Table 2 highlights several features of the program's development in terms of the three-term interactional model. Along the top are the three interaction components.

TABLE 2

Antecedent Events (S^D)	*Arithmetic Behavior* (R)	*Consequent Events* (S^R)
1. "Say sixteen."	"sixteen"	"Good work" plus tokens on CRF.
2. "Count from 18 to 20."	"Eighteen, nineteen, twenty"	"Nice counting" plus tokens.
3. ⋮ ⋮	⋮	⋮
39. A written story problem.	Read problems.	Altered written problems.
Altered written problems.	Translation of numbers and operations into computation format.	Problem in computational format.
Problem in computation format.	Computation chains and solution responses.	"Nice job, good work, etc.," on VR 10.

The numbers on the left represent program sequences, that is, sections of the program designed to teach components of the terminal response. Sequence 1 is the entering behavior or skill level of the children on enrollment, verbal imitation of numbers. Sequence 39, at the bottom, shows an analysis of a terminal story problem. As the child moves through the program, the antecedent stimuli increase in complexity and form from a verbal "sixteen" to a written problem. The behavior also increases in complexity. While the terminal solution response may still be sixteen, the chain leading to the solution has expanded in length and breadth. In the consequences column, reinforcement systematically decreases throughout the program from continuous social and token reinforcement to only occasionally scheduled social reinforcement. The entire interactional sequence can be viewed as the development of arithmetic behavior and independence. That is to say, as the child progresses through the program, he can do more complex arithmetic problems with less teacher intervention.

The initial development of the program proceeded from the terminal behavior back to the point where the children could enter and be successful. A task analysis was made of the terminal interaction (story problems) which isolated component arithmetic responses necessary to fulfill the final objective. These behaviors were ordered into an instructional sequence so that as the child progressed through the program he would master the most basic skills first, and subsequent behaviors would build upon the basic interactions as well as maintain them in the child's repertoire. For example, in row 2 is a segment of the second sequence, number chaining or counting. Obviously saying the numbers is prerequisite to saying numbers in a set order. Sequence 1 is precurrent to task 2, and task 2 builds upon and maintains the behaviors mastered in the first program unit. Such an analysis was carried out from sequence 39 to sequence 1.

The sequence of the actual program is shown in Table 3. For convenience, it has been grouped into three divisions: (1) precomputational arithmetic, instruction on prerequisite behaviors for addition and

TABLE 3

SEQUENCE OF STIMULUS ITEMS

Ⓐ Precomputational Arithmetic Skills

1. Verbal chaining of numbers
2. Counting number (sets)
3. Discrimination of written numbers
4. Naming written numbers

Ⓑ Simple Computational Arithmetic Skills

1. Finger counting for addition & subtraction
2. Introduction of zero
3. Double column

Ⓒ Arithmetic Comprehension

1. Story problems (ties in with reading, writing and spelling programs)
    a. Translation of stories into computational format for solution; both from auditory and written stories.

subtraction; (2) computational arithmetic, instruction on addition and subtraction; and (3) arithmetic comprehension, instruction specific to story problems. Since space limitations make it impossible to present all 39 sequences which make up these units, several selected units and monitoring data on four children are presented and discussed.

PRECOMPUTATIONAL ARITHMETIC

Precomputational arithmetic begins with instruction on counting for a child with an imitative repertoire. Figure 1 shows monitoring data on the counting sequence for once child, J. C. Intermittent reliabil-

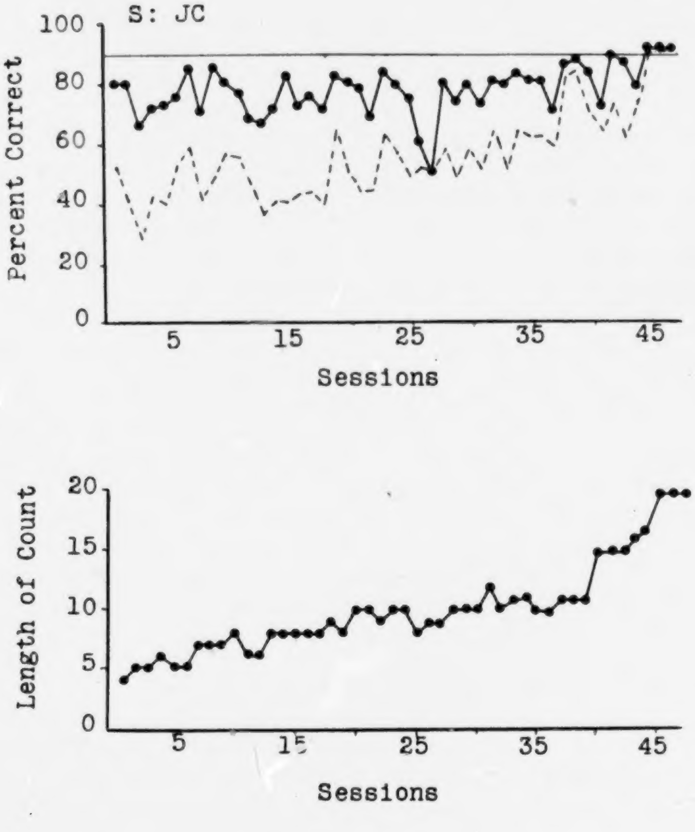


Fig. 1. Monitoring data on the counting sequence for one child, J. C. In the top figure (counting accuracy), the upper curve, solid line, indicates overall accuracy; the lower curve, broken line, indicates unprompted accuracy. The lower graph shows progression through the counting sequence in terms of chain length.

ity of scoring was accomplished with the aid of a trained observer. Comparisons of counting accuracy, prompts, and reinforcement schedule yielded agreements on all measures above 90%. The upper curve, solid line, is the child's overall accuracy scores. The lower curve, broken line, indicates accuracy when prompted frames are scored as errors. Prompted frames are those in which the tutor gives hints to the answer. For example, in the counting sequence, the tutor may prompt the response "twenty" by saying "twa." Prompts, then, are tutor-added supplements to the antecedent stimulus. The difference between these two curves indicates the extent the child is dependent on the tutor's help. While the child's overall accuracy varied between 70% and 90%, his unprompted accuracy showed a rather steady increase to the 90% mastery criterion in the lower curve. The lower graph indicates progression through the counting sequence. Longer chains were required for reinforcement as the child mastered shorter counts. The child became less dependent as he progressed through the program and his accuracy increased.

The daily monitoring of the two accuracy measures was suggested by the programming principle that pupils should make minimal errors; hence, be successful and reinforced a high proportion of the time (Skinner, 1968). The teacher can scrutinize these two data and determine the amount of prompting that is necessary to keep the child successful and advancing toward independent behavior.

The remainder of the precomputation division of the program deals with counting objects, matching numbers to numbers, and naming numbers. Example frames are shown in Figure 2. The three frames at the top under *Numbers and Sets* are examples of exercises which require the child to match sets of objects, first similar, then objects differing in form and of increasing number. The middle two frames under *Number Discrimination* are examples of items which require the child to match written numbers. The last three frames under *Number Naming* are examples which deal with number naming. These frames combine counting and matching to prompt identification of numbers.

COMPUTATIONAL ARITHMETIC

The second division of the program--simple computational arithmetic skills--is concerned with instruction on addition and subtraction. Figure 3 shows data on the acquisition of addition and subtraction for one child, S. M. The top curves are overall accuracy scores. The lower curves are unprompted accuracy scores. Here, again, prompted frames are scored as errors. Unprompted accuracy curves are a stringent measure of unit mastery. The first pair of curves under _A_ show rapid acquisition of a finger-counting procedure for addition. The second pair of curves *(B)* indicates expansion of addition to where both numbers varied from 1-19 with sums no greater than 20. The third set of curves, under _C_, shows acquisition of sub-

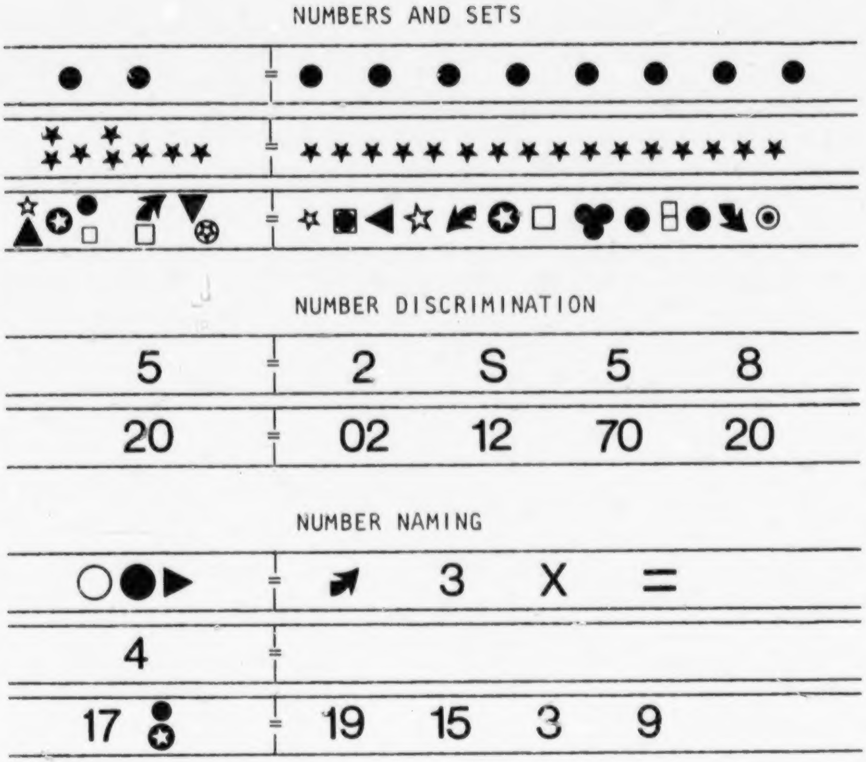


Fig. 2. Example frames from precomputational arithmetic: the top three frames deal with numbers and sets; the middle two frames, number discrimination; and the bottom three frames deal with number naming.

traction behavior and maintenance of addition. Up to this point in the program the child was selecting answers in a multiple-choice format, circling written answers. The last pair of curves *(D)* shows performance on mixed addition and subtraction where the child was required to construct, or write his answer. The rapid mastery of this unit attests to the mastery of prerequisite arithmetic and writing skills. These data indicate two things. First, they indicate the beneficial role of prompting. Complex chains of responses (addition and subtraction) were quickly developed by maintaining a high rate of success; hence, reinforcement (i.e., high overall accuracy). Prompts were faded depending on the data until each unit was mastered to the 90% criterion (i.e., high unprompted accuracy). Specifically, when overall accuracy was above 85%, prompts were systematically deleted. If overall accuracy fell below 85%, we would temporarily back up, adding prompts to increase the child's success. Second, these data indicate the effectiveness of the program. The lower set of curves is an accurate measure of the acquisition of independent problem solving. Teaching this child to add and subtract independently required twenty-five hours of instruction.

The remainder of the computation division involves instruction on addition and subtraction in a vertical format, the use of zero, and several units designed to increase the complexity of the problems to double columns and multiple rows.

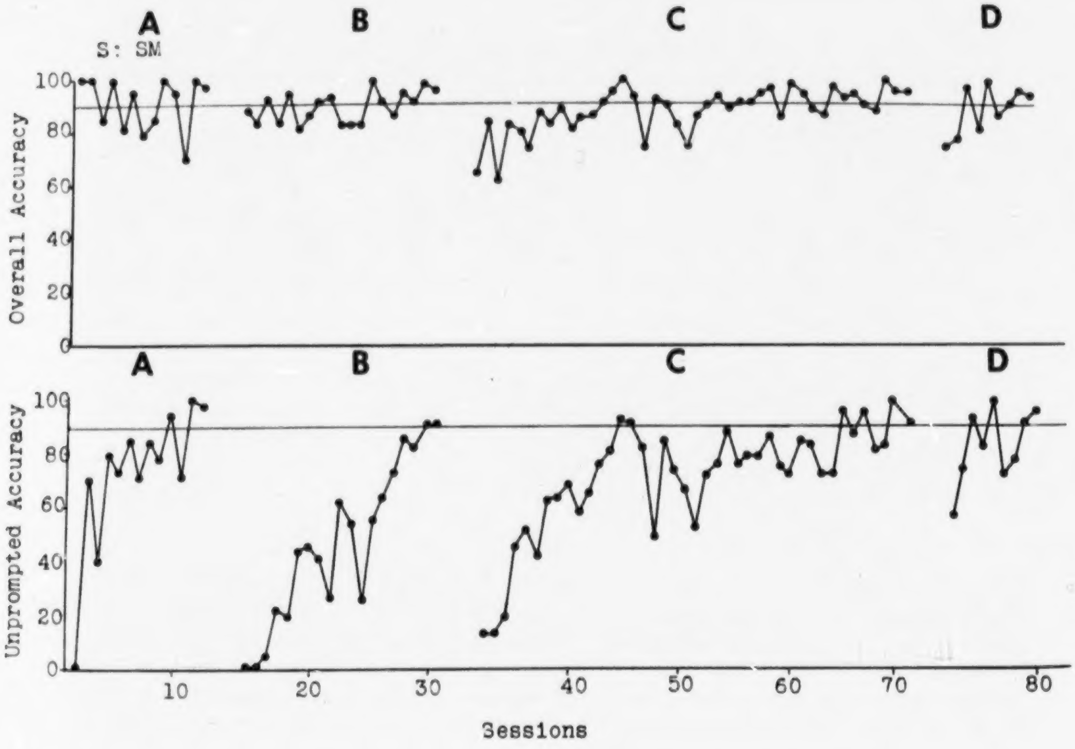


Fig. 3. One child's overall accuracy (top graph) and unprompted accuracy (bottom graph) for four units of instruction on addition and subtraction.

Basic Research in the Context of Computational Program

The following study compared the effects of two attending programs on the acquisition and maintenance of double-column mixed addition and subtraction. This is an example of basic research within the classroom setting. Figure 4 shows characteristics of the two attending programs and three problem types. Attending here, refers to some overt response to the + and - operation signs. PROGRAM A reinforced the child for circling the plus or minus sign *prior to doing each problem.* The attending response was the same for + and -, a circle, and was the same for beginning as well as terminal frames. The beginning frames in PROGRAM A were of type 1 and 2. It is obvious, even with the sign deleted, that problem type 1 is an addition problem because the upper

ATTENDING PROGRAMS

	BEGINNING FRAMES	TERMINAL FRAMES
PROGRAM A	CIRCLING + and - TYPES 1 and 2	CIRCLING + and - TYPE 3 only
PROGRAM B	CIRCLING and NAMING + and - TYPE 3 only	CIRCLING + and - TYPE 3 only

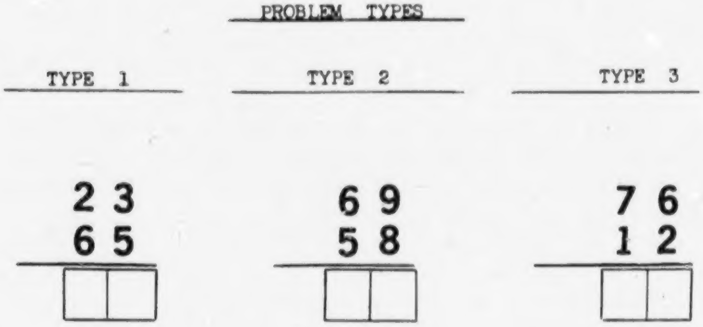


Fig. 4. Description of conditions for Attending Programs A and B (top) and example problems of Type 1, 2, and 3.

numbers are smaller than the lower numbers. Type 2 problems are less revealing, but the children have not yet learned to carry and have been reinforced for writing only one digit per answer box. With these two restrictions, it is obviously a subtraction problem, even with the operation sign absent. Therefore, type 1 and 2 problems could be solved without paying attention to the signs present. The terminal problems in PROGRAM A were type 3 which require close inspection of the sign for accurate solution. Ignoring sign, they could be solved either by addition or subtraction without violating the child's two "rules." Circling the signs was still required on terminal type 3 frames.

PROGRAM B also reinforced the child for overt circling of the signs, but beginning problems also required the child to name the operation sign, plus or minus, as well as circle it. That is, PROGRAM B required the child to make two different responses to the signs while PROGRAM A required the same response. Problems were always of type 3 in PROGRAM B. Thus, the terminal items for both PROGRAMS A and B were identical, type 3, requiring circling only.

Data presented are on terminal items. Figure 5 shows one child's performance on these terminal items following initial histories on beginning frames of PROGRAMS A and B. After going through the beginning of PROGRAM A, where only circling of signs was reinforced, the child's accuracy was low, averaging about 70%. All errors were due to performing the wrong operation. This was true even though on previous units of the program she had 100% accuracy on addition and subtraction. In between the first two curves the child completed PROGRAM B. The child was reinforced for naming the sign for several sessions then switched to the terminal task, circling only. It can be seen the program was effective. It was only necessary to initially reinforce different attending responses to the sign to increase accuracy to criterion, and attending was thereafter maintained with only a circling requirement. The next two curves are a replication. Again, completing PROGRAM A deteriorated the child's performance and the subsequent requirement to differentially name the signs in PROGRAM B remediated the problem.

These results suggest that for attending to be optimal, the child must make more than an overt response; he must respond differentially to the materials. Reinforcing correct solutions was not enough to increase attending once it deteriorated (PROGRAM A - terminal problems). However, reinforcing correct solutions did maintain efficient attending once the child had responded differently to the two operation signs (PROGRAM B). Findings from this study were immediately applied. The program was redesigned to eliminate problem types 1 and 2, and include the differential naming response to the operation signs. Subsequent children going through the sequence have been successful.

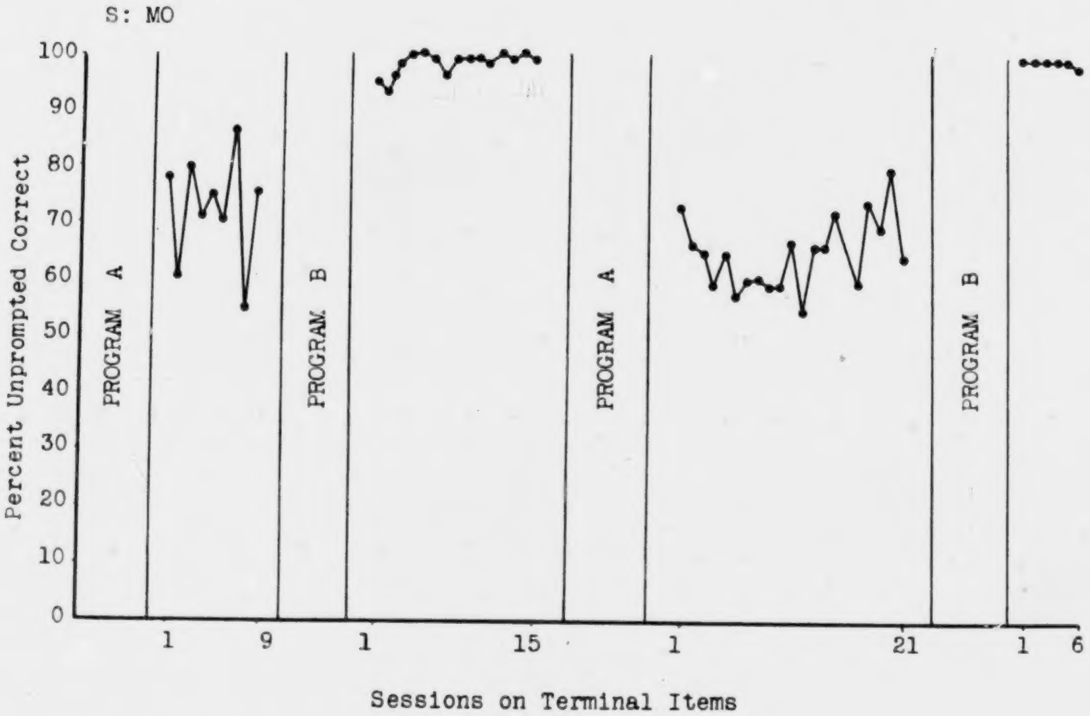


Fig. 5. One child's accuracy performance on terminal frames following histories of Attending Programs A and B.

ARITHMETIC COMPREHENSION

The third division--arithmetic comprehension--brings computation under the control of written sentences. Figure 6 shows a shortened sequence of this part of the program. Visual prompts are provided in the materials and are gradually faded. Initially, boxes are provided for each digit needed as well as the operation sign. Also, the digits required in the answer boxes are in the sentence just adjacent to the computation area. At the same time these physical prompts are being faded, the problems increase in complexity. Verbs expressing operation are expanded (e.g., "put" and "take" to "finds" and "forgets"), numbers are written out, sentence length is increased, and the order of sentences is made more complex.

SAMPLE STORY PROBLEM TYPES

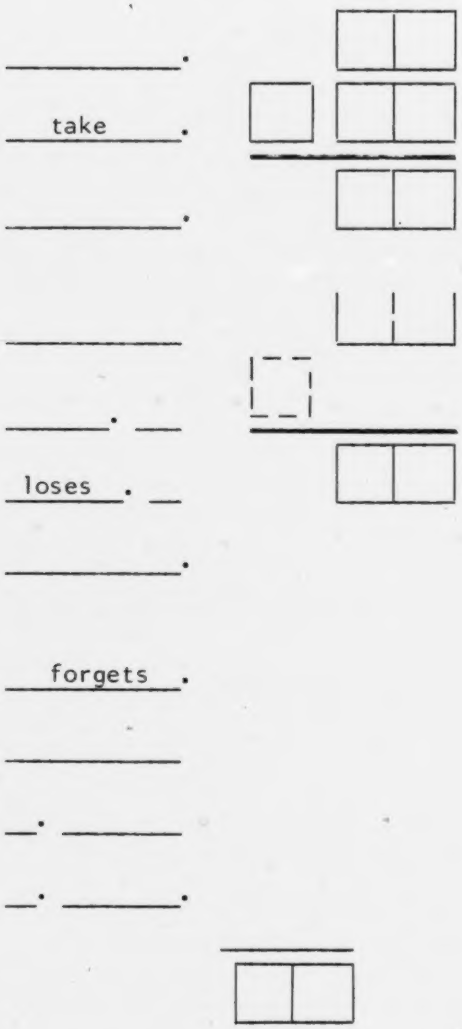

take
loses
forgets

Fig. 6. Story problem frame format from the arithmetic comprehension program.

Figure 7 represents one child's unprompted accuracy across daily sessions on the story problem sequence. With the exception of a few points, the sequence was optimal. On only several occasions was it necessary for the tutor to prompt this child. This girl began the year unable to reliably count from 1-20, but nevertheless was able to master complex story problems after eight months of instruction.

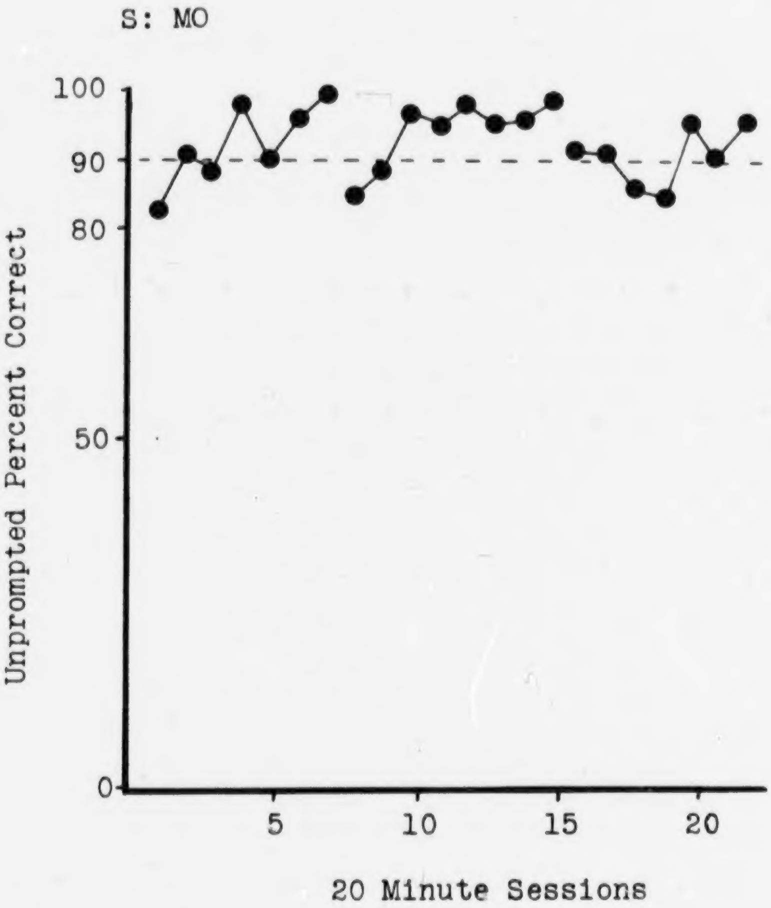


Fig. 7. One child's unprompted accuracy performance on the arithmetic comprehension program.

SUMMARY

A primary arithmetic program is under development using behavior principles. Arithmetic behavior was analyzed as verbal behavior involving the three-term contingency. The unique features of arithmetic noted were:

1. Arithmetic involves chaining simple responses into long sequences, and bringing these responses under the joint stimulus control of the numbers and operation signs.

2. Arithmetic is a logical system with great potential for consistent reinforcement. The fact that solutions are easily checked by the child suggests methods to establish being correct as a reinforcer may be fruitful.

3. In order to make arithmetic less boring, and provide maintenance for learned behaviors, it should be incorporated into the child's daily life. This includes using arithmetic in school settings which provide a wider range of reinforcement than arithmetic alone (e.g., sciences), and in the child's non-school habitat (e.g., buying goods and services).

4. The program was developed to teach only those behaviors necessary to master the terminal objectives, and to teach these component behaviors in a sequence where each succeeding behavior built on interactions already mastered. Prompts were added, as needed for individual children, to keep reinforcement high.

5. The data presented suggested how daily records were important in monitoring the children's progress and for making alterations in the program.

The programming principles involved in the development and administration of this program are equally applicable to more complex arithmetic behaviors. In fact, when children have larger behavior repertoires, it is easier to vary materials and make computation more functional for obtaining naturally occurring reinforcers.

REFERENCES

Skinner, B. F. *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.

Skinner, B. F. *The technology of teaching*. New York: Appleton-Century-Crofts, 1968.

A FINAL NOTE

j.a. grimm

I would like to tell a story, the theme of which many readers are probably familiar. When I was a freshman in college, I worked in a summer camp for physically handicapped and retarded children. I soon found that I was not expected, encouraged, or able to teach these children anything. My sole function was to keep them clean, warm, and comfortable. So, naturally, I became a psychology major. As a graduate student, I had occasion to visit a state training school for retarded children. I had, in the meantime, encountered companies of rats in underwater mazes, battalions of college sophomores in front of memory drums, and armies of clinical students hotly debating the merits of eclectic therapy. But, I found that the staff at the training school was not expected, encouraged, or able to teach the children anything. The staff's major function was to keep the children clean, warm, and comfortable.

Bijou (1970) has posed the question, "What does psychology have to offer education - now?" I think that the answer to that question is becoming increasingly clear. It is a concentration on behavior/environment interactions. It is a body of techniques for managing the setting, antecedent, and consequent events that comprise the child's learning environment. It is an optimistic attitude that the retarded child can learn. But it is also hard work and very demanding of teachers. Teachers should not be misled into the belief that the behaviors required of them are easy.

When plans for the Child Behavior Laboratory were formulated in 1965, curriculum materials that were both suitable for young retarded children and commercially available were non-existent. Today, most materials for this population still suffer from two defects. First, there is little available evaluative data concerning the remedial effect of the material on the child. Second, available materials do not provide the teacher with sufficient procedural detail for their effective administration. The research presented in the preceding papers has been concerned with the development of both materials and procedures and with the evaluation of these products.

Although the task is still incomplete, results to date indicate that psychology has a great deal more to offer education than a technology for keeping retarded children clean, warm, and comfortable. A brief case history of a five year old girl named Laurie is illustrative. She was diagnosed as brain damaged and retarded, and eliminated from kindergarten for "failure to learn anything." Indeed, Laurie could not count from 1 to 20, could not copy any letters, could not read any words (in fact, she could not even discriminate one group of letters from another), and responded emotionally (i.e., cried) to most academic tasks when she was first enrolled in the Laboratory class. As a consequence of the academic programs and procedures presented in the preceding papers, Laurie learned to add and subtract, mastered manuscript writing, acquired a fifty word spelling repertoire and a sixty word reading vocabulary. She also learned to work accurately for long periods without contrived reinforcers and with only intermittent teacher attention. She will be placed in a regular public school first grade classroom in the coming school year. Laurie, then, is representative of what psychology has to offer education now.

REFERENCES

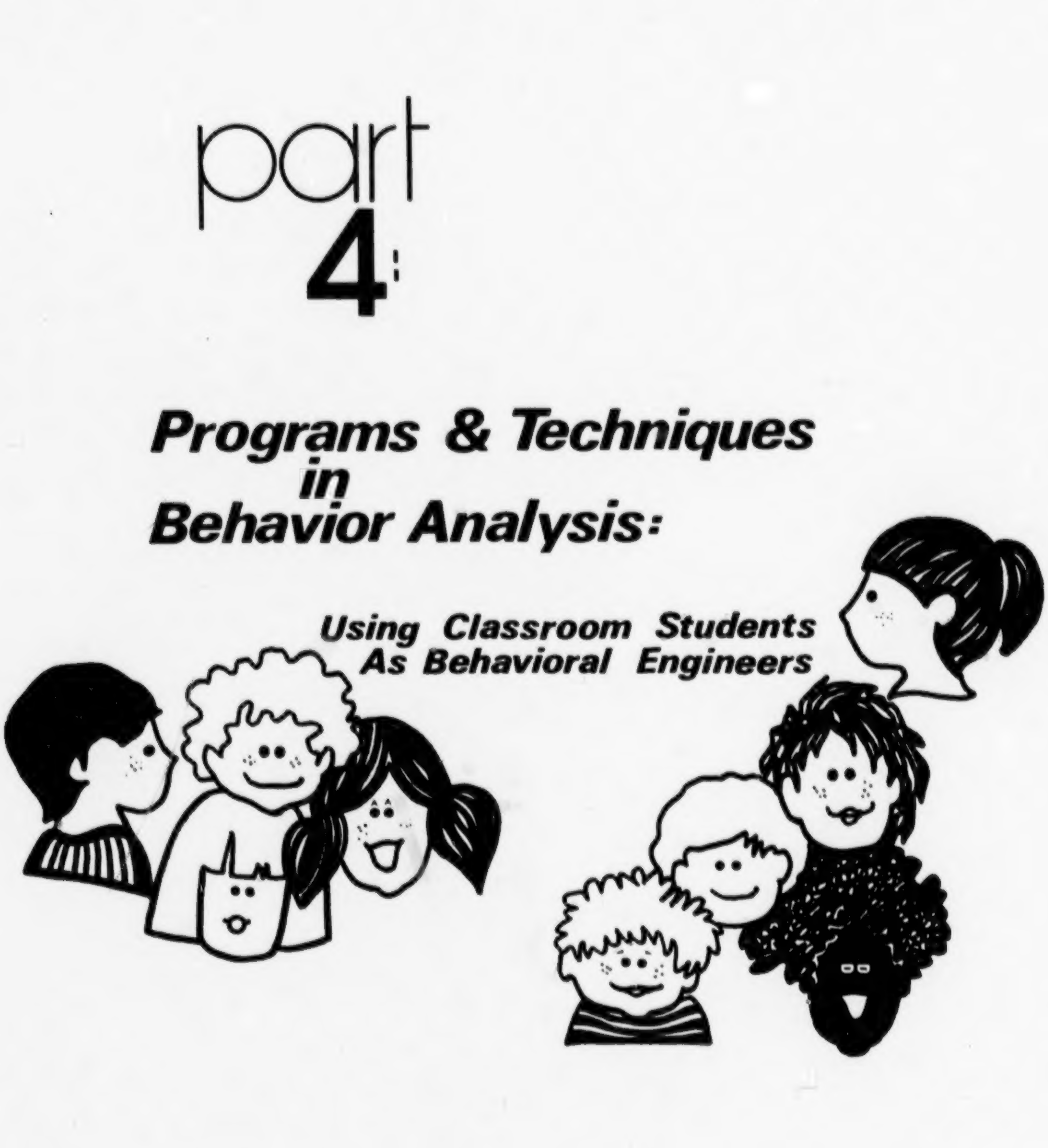
Bijou, S. W. What psychology has to offer education - now. *Journal of Applied Behavior Analysis*, 1970, 3, 65-71.

* * *

part 4

Programs & Techniques in Behavior Analysis:

Using Classroom Students As Behavioral Engineers



THE EFFECTS OF A PEER CORRECTION PROCEDURE ON THE ARITHMETIC ACCURACY FOR TWO ELEMENTARY SCHOOL CHILDREN

miriam f. conlon
carolyn hall
edward m. hanley

This study explored the effects of a peer correction procedure on the arithmetic accuracy of two children with academic deficits. The ABAB experimental design included assessment of the peer corrector's daily performance. During the peer correction conditions, when the peer systematically marked big, red C's by correct responses and ignored incorrect responses, each pupil's accuracy increased when compared with Baseline conditions. No change was noted in the peer's accuracy during each condition. This study indicates the effectiveness of using school age children to provide positive consequences for academic accuracy.

Peer influence has been demonstrated to be an effective technique for modifying behavior in the classroom; e.g., social behavior (Wolf, Hanley, King, Lachowicz and Giles, 1970) and academic behavior (Evans and Oswalt, 1968; Burdett, 1969). In spite of successful changes reported in the behaviors of the subjects in all studies there is no measurement of the manipulator's behavior during any phase of the studies.

In the present study, measurements were taken not only of the daily academic work of the tutor's students, but of the tutor's daily work as well.

Subjects and Setting

Three subjects were selected for the experiment on the basis of their scores in a series of pretests on arithmetic review examples. The pretest was given to several students; then one student was chosen as peer tutor or corrector on the basis of a generally high but variable score. The two other students were chosen as tutees on the basis of low, but variable scores.

The classroom contained eleven students between the ages of eight and ten. The students were placed in this room because they were below grade level in math or reading or both.

The corrector was Tom, who had been in an ungraded classroom since he began school because of lack of readiness originally and then because of poor achievement. He also seemed to lack self-control and, therefore, was a constant discipline problem. He was in a low grade three reader and working on grade level in math.

Dick, age 9, had been in an ungraded situation for one year, in a regular classroom for two years, then placed again in an ungraded situation because of a very low reading level. He was reading on a second grade reading level and on a fourth grade level in math.

Harry, age 9, had been in an ungraded situation for one year, in a regular classroom for two years, then placed again in an ungraded situation because of reading comprehension deficits. He was reading on a second grade level and working on a fourth grade level in math.

Each afternoon the whole class was given a sheet of paper containing thirty review math examples similar to those on the pretest. The sheets were part of a series arranged as a review exercise but of increasing difficulty as new facts became learned facts and then review facts. The teacher instructed the students to complete as many as possible within the 20-minute period.

PROCEDURES

Baseline 1

The first six days was a Baseline period during which the review sheets were handed out and the instructions were given by the teacher as usual. The only change that was made was that a permanent seating arrangement was established. This included seating the three subjects at the same table so that when the first experimental condition was introduced, additional confusion would be minimal.

Peer Corrector I

During experimental condition I, the teacher gave a stopwatch and an answer sheet to the corrector, Tom, and explained the procedure to him and the two other subjects. At the end of six minutes Tom stopped the watch and, using the answer sheet, corrected the examples completed by Dick and Harry. This involved marking a C with a red pencil on the correct examples and ignoring incorrect or incomplete work. Tom then returned to his seat and resumed work. Dick and Harry could continue work except when Tom was correcting *his* paper. The watch was again set in motion and when six minutes had elapsed the procedure would be repeated. If time remained, Tom would correct for the last time at the end of the 20-minute period.

Two days after the experimental condition was in effect, Harry asked if he could operate the stopwatch at a later period in the day. This request was granted, contingent upon a 100% correct score on the math paper.

Baseline 2

To test the strength of the experimental conditions, the classroom teacher removed the contingencies on day 18 and math period was returned to Baseline conditions for four days.

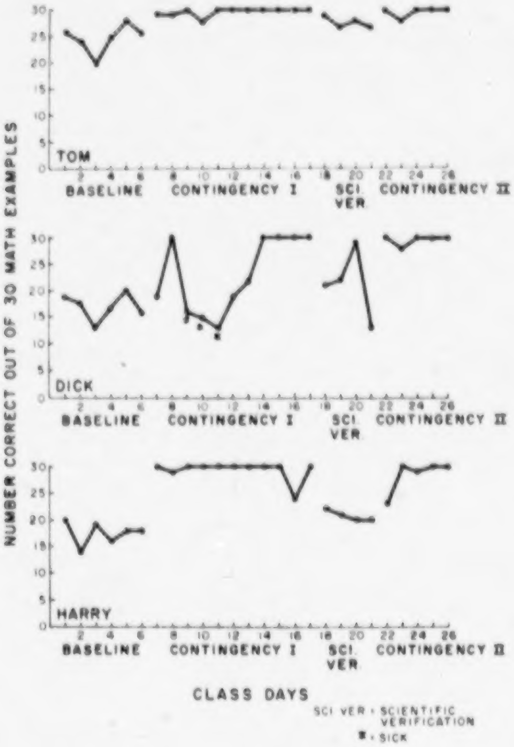


Fig. 1. The teacher's daily record of each pupil's accuracy on arithmetic review examples during each experimental condition. Tom (upper graph) was the peer corrector, while Dick (middle graph) and Harry (bottom graph) were the tutees.

Peer Correction II

Because the experimenter was primarily concerned with effects of a peer tutor (corrector) on the math scores of the tutor and his subjects, it was decided to eliminate use of the stopwatch, contingent on 100% correct scores, by anyone other than Tom who continued with this same procedure. On day 22, the teacher explained the difference between the two contingency periods to the class and class was conducted under these conditions for the remainder of the school year. During all conditions the teacher corrected Tom's paper and simply checked corrections made by Tom during the experimental conditions.

A second person occasionally observed the whole 20-minute period, and checked the subjects' papers.

RESULTS

As Figure 1 shows, accuracy on arithmetic worksheets improved for all three subjects during both experimental conditions as compared with accuracy during Baselines 1 and 2.

	TOM	AVERAGES FOR P^1	
B-1	20-28 correct or *67%-93%*	B-1	*83%*
PC-1	28-30 correct or *93%-100%*	PC-1	*99%*
B-2	27-29 correct or *90%-97%*	B-2	*93%*
PC-2	28-30 correct or *93%-100%*	PC-2	*99%*

	DICK	AVERAGES FOR P^2	
B-1	13-20 correct or *43%-67%*	B-1	*57%*
PC-1	13-30 correct or *43%-100%*	PC-1	*84%*
B-2	13-29 correct or *43%-97%*	B-2	*73%*
PC-2	28-30 correct or *93%-100%*	PC-2	*99%*

	HARRY	AVERAGES FOR P^3	
B-1	14-20 correct or *47%-67%*	B-1	*58%*
PC-1	24-30 correct or *80%-100%*	PC-1	*98%*
B-2	20-22 correct or *67%-73%*	B-2	*69%*
PC-2	23-30 correct or *77%-100%*	PC-2	*95%*

On days 10 and 11 Dick had been absent in the morning because of illness and on day 12 he was dismissed early, when he complained of a headache.

DISCUSSION

The results of this study demonstrate that a student can work successfully as a corrector with peers with no adverse effects on his own work. In fact, the tutor's *percent correct* average improved during the experimental conditions concurrently with his students.

The first two days of the study required continuous monitoring by the experimenter to insure correct use of both the stopwatch and the answer sheet. The answer sheet proved to be no problem as it was turned face down except during the correcting times and the stopwatch was appropriately operated by day 3. Once the *trouble shooting* was completed, occasional checks by the experimenter and daily reports from the classroom teacher were ample evidence that the experiment could be conducted independent of adult supervision. Tom, the corrector, ceased to be a discipline problem, at least during the math period. Towards the end of the experiment he would demonstrate his arithmetic competency by correcting the papers without using the answer sheet.

REFERENCES

Burdett, C. A peer tutor study. In McKenzie, H. S., *The 1968-1969 Report of the Consulting Teacher Program: Volume II*. Burlington, Vermont: Special Education Program, College of Education, University of Vermont, 1969, pp. 206-209.

Evans, G. and Oswalt, G. Acceleration of academic progress through the manipulation of peer influence. *Behavior Research and Therapy*, 1968, 6, 189-195.

Wolf, M. M., Hanley, E. M., King, L. A., Lachowicz, J., and Giles, D. K. The timer game: a variable interval contingency for the management of out-of-seat behavior. *Exceptional Children*, 1970, 37, 113-116.

A BEHAVIORAL APPROACH TO REMEDIAL READING USING STUDENTS AS BEHAVIORAL ENGINEERS[1]

jerry willis
jeane crowder
betty morris

The problem of children who read poorly has probably received as much attention as any other problem in education. Models for special programs to remediate deficits abound. Many educators advocate correcting some underlying problem of which poor reading is assumed to be a symptom. Some advocate an emphasis on perceptual training (Silver, Hagin and Hersh, 1967); some use intensive phonetic practice to improve reading (Tomatis, 1969), and others provide linguistic remediation (Hayes, 1967) to improve reading.

Recently, successful behavioral approaches to reading have also been described in the literature (Willis, Morris and Crowder, 1972; Staats, Minke and Butts, 1970; Gray, Baker and Stancyk, 1969). Generally, the behavioral approaches focus on shaping the target behavior, reading, and provide systematic consequences for appropriate reading behavior. A frequent question concerning the use of behavioral reading methods relates to their effectiveness in the absence of consideration of *underlying causes* such as a visual perceptual problem which interferes with reading. To deal with this question the current study was carried out in a poverty area school, used a behavioral approach, accepted *any* student referred who had a reading problem, and made no attempt to diagnose or remediate any problem other than a reading deficit. The program was supervised by the school counselor and managed by a paraprofessional hired by the school system. Two psychologists from the School Consultation Project served as consultants. The training began in October and ended in April.

Subjects

Ss were 43 children in grades 2 through 8 who read two or more grade levels below their grade placement or who were judged by their teachers to have serious reading deficits (e.g., a second grader who read at a first grade, second month level). These children all attended the same school in Birmingham, Alabama and would be labeled culturally deprived,

---

[1] The authors would like to express their appreciation to Mr. James Moreno, Miss Margaret Loranz, and Mrs. Mariah Poellnitz.

as having a low socio-economic level, or poor. The mean I.Q. (Slosson Intelligence Test) was 82. Although about 75 children worked in the program over the year only 43 remained at the school from the pretest in October to the posttest in April.

Setting

The school provided a regular classroom for the project. Over the year the eighth grade Behavioral Engineers were encouraged to modify the room as they wished. Bulletin boards, posters, and displays were made, and a rather dingy, dirty room was turned into a bright, clean work area. Children came to the room for thirty minutes of training each day.

Method

The program drew from several of the behavioral approaches referenced above. Students referred by teachers were given a short individual I.Q. test (Slosson) and a short reading test (Slosson Oral Reading Test) by the school counselor. However, no students were eliminated because of low I.Q. or extremely low reading scores. The remedial program itself had four major components:

★ MATERIAL

Interesting reading material individually prescribed for each student was available. Every child had material that was both within his reading ability and interesting. Children who were non-readers began with the *Rebus reading program* (American Guidance Services, Publishers Building, Circle Pines, Minnesota 55014). Those with minimal skills entered the *Sullivan Programmed Reading Series* (Webster Division, McGraw-Hill), while students with slightly better reading skills were assigned to the *SRA Reading Lab* (SRA, 259 E. Erie Street, Chicago, Illinois 60611), the *Know Your World* and *You and Your World* newspapers (American Education Publications, Education Center, Columbus, Ohio 43216) or the *Readers Digest Reading Series* (Readers Digest Services, Inc., Pleasantville, New York 10570). However, there are many "high-interest, low vocabulary" books that would have been equally appropriate.

★ STAFF

Twenty-three eighth grade students actually carried out the training. The only requirement for selection was that the student wished to participate and had a free period during the day. The eighth graders (Behavioral Engineers) received five hours of training in principles of behavior modification as they apply to reading. Role playing, modeling, and behavior rehearsal were used extensively. The Behavioral Engineers (BE) divided themselves into five teams and selected names such as *Mod Squad* and *Fantastic Four*. Each team had its own time during the day when its members worked. The BE were regularly observed by the school counselor who used a checklist of 20 aspects of the remedial program (see Appendix A). The team with the highest "efficiency rating" received a large trophy which was displayed in their section of the room. Second and third place trophies were

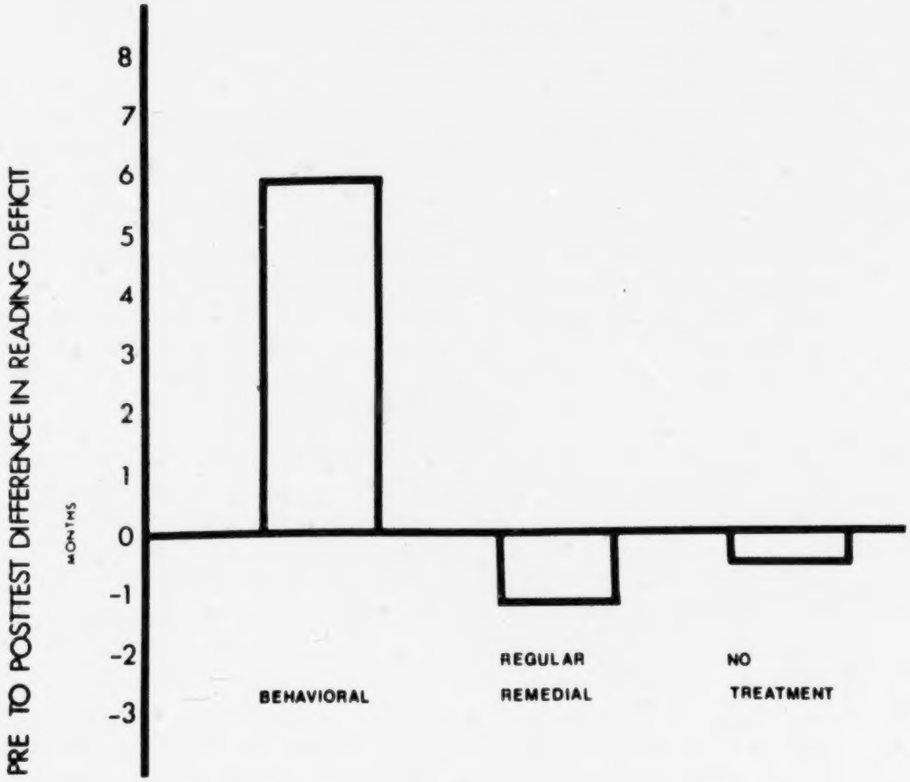


Fig. 1. Pre- to post-treatment differences in reading deficit for the children taught by Behavioral Engineers (8th grade students), those in a regular remedial program, and those who received no reading instruction outside their regular classroom. Students in the behavioral reading program reduced their reading deficit while students in the other two conditions were actually further behind at the end of the 7-month training program.

also given so that only the two lowest teams received no recognition. Praise for good work was regularly administered to the BE by the counselor and psychologists. Twice during the year movies of the BE were made and shown to their homerooms in an effort to give them recognition and status for excellent work.

Although the BE worked with children from grades two through eight for the first week, the paraprofessional who managed the program worked with the 7th and 8th graders for the remainder of the time when relationships between the BE and children their own age appeared less positive than with younger children.

TUTORING TECHNIQUE

The third component was a special remedial reading technique that emphasized immediate rewards for successful reading. Each BE worked with a pair of students of like ability and reading level. As a student read, the BE provided a green plastic chip (American Guidance Services) for each sentence of five words or more read correctly. (In a few instances children were rewarded for each correct word because of a very low rate of correct reading.) The BE also was taught to pair praise with the green chips and to show enthusiasm for successful reading behavior. When a student made an error the BE corrected the error and helped the student finish the sentence. Then he gave the student a red chip, and the other member of the pair was allowed to read and receive chips until he made an error. The student began with the sentence in which the error was made, thus reinforcing attending while the other student read. If a student read ten sentences in a row correctly the reading task was automatically transferred to the other reader.

DATA COLLECTION

The final component was the daily collection of data. Each pair of readers counted his red and green chips at the end of the period and recorded the number on a record sheet. Praise was liberally given for progress. Competition was encouraged between readers in matched pairs but not across pairs.

Another major use of the data was in planning future work. If a pair of readers had an error rate near zero for several days, the BE moved them into more difficult material. On the other hand, if the daily data reflected a high error rate, the pair was placed in less difficult material. If one reader consistently made the highest score the pair was re-matched.

Two pairs were broken up and taught on a one-to-one basis when the data indicated the competitive condition was not effective for them.

RESULTS

Table 1 presents mean pre and posttest data for the 43 children in the Behavioral Reading Program. These data were evaluated using a technique described by Libaw, Berres and Coleman (1965). Pretest reading scores were used to predict the progress of students during the 7-month program. Pretest reading scores ranged from 0.0 to 3 years, 6 months. A child who had made a year's progress in two years of schooling was expected to gain three months in the next six months. Of 43 students, only three failed to exceed their previous rate of progress, a significant improvement (Sign Test $x = 3$, $p < .00003$).

TABLE 1

MEAN TEST DATA ON THE BEHAVIORAL READING GROUP.

TEST	PRETEST	POSTTEST	DIFFERENCE
*SLOSSON INTELLIGENCE TEST	*82*		
*SLOSSON ORAL READING TEST	*1.8 years*	*2.8 years*	*+1.0 years*

Since the Slosson Oral Reading Test is basically a test of word recognition, there was a possibility that the behavioral reading program improves word calling skills without improving comprehension. To test this possibility the first 23 students tested for the Behavioral Reading Group were given the Peabody Individual Achievement Test, Reading Recognition and Reading Comprehension subtests. These students made a mean gain of 6.5 months on the Reading Recognition subtest and 6.3 months on the Reading Comprehension subtest indicating equivalent gains in both areas.

The Behavioral Reading (BR) Program was also compared to a Regular Remedial Reading (RRR) Program staffed by a certified, degree holding teacher. Students attended this program for thirty minutes a day in small groups of about ten. Originally, it was planned to compare the total BR group to the total RRR group of about fifty students. Pretest data, however, revealed that the RRR group was composed primarily of children with mild reading problems or no problem at all while the BR group was made up of children with more serious reading deficits. In fact, many of the children in the BR group were "rejects" from the RRR group, children considered too far behind or too stupid to profit from special help. The reluctance to accept children with serious reading problems was encouraged by a system-wide policy that children more than two years behind could not profit from the remedial reading program. The mismatch was further compounded by the experimenters' willingness to take anyone into the BR group. Forty-five percent of the BR group had I.Q.'s in the educable mentally retarded range (80 and below). For research purposes, thirteen students in the RRR group were matched as closely as possible with thirteen BR students. Table 2 presents the pre- and post-reading test data for the matched groups.

TABLE 2

MEAN GRADE LEVEL AND TEST DATA FOR THE THREE GROUPS.

GROUP	GRADE	I.Q.	PRE-SORT	POST-SORT	DIFFERENCE
Behavioral N = 13	4.4	79	2.70	4.00	1.30
Regular Remedial N = 13	4.4	89	2.98	3.40	.58
No Treatment N = 6	6.1	75	4.25	4.90	.65

Also included is a group of six students who were tested by the counselor but who were in no remedial program. Since this group was not matched, comparisons with the matched groups must be considered tentative.

Using the Sign Test the BR and RRR groups were not significantly different in grade placement, I.Q., or pre-reading test scores. However, the BR group did significantly better (Sign Test $x = 2$, $p < .046$) on the posttest, and its gain of 1 year and 3 months was significantly better (Sign Test $x = 2$, $p < .019$) than the 6-month gain of the RRR group.

In comparing the two groups it should be remembered that any remedial program must, by definition, help correct the problem which brought the program into existence. To eliminate a reading deficit a remedial program cannot merely produce a gain in reading. A 6-month program which produces a 6-month gain merely leaves the student no further behind than he was when he started. To be effective a remedial program must produce progress at an above average rate if the student is ever to overcome his reading deficit. Figure 1 presents the scores in the form of changes in the reading deficit of the child. The BR group made the seven months progress expected of "average" children and an additional six months progress on their reading deficit. The RRR group was actually slightly further behind at the end of the program as was the group which received no special program.

DISCUSSION

Even when the relatively crude pre to posttest data are used, the behavioral reading approach was markedly superior to the regular remedial reading program. Further, a cost analysis indicated the BR program cost about $18 a day while the RRR approach cost the school system about $45 a day for a comparable number of children. Money alone does not automatically produce a better learning environment.

The success of BR is perhaps as important for what it did not do as for what it did. BR made no attempt to provide compensatory training in *underlying causes* such as visual-motor deficits, psycholinguistic liabilities or gross motor patterning; no expensive or exotic materials were used; training of the teachers lasted about five hours rather than requiring a year or more of graduate training in education, and finally, no tangible reinforcers such as toys or candy were used.

However, the BR program was not without frustrations to the staff. A problem which quickly became apparent was the interest and sensitivity of the Behavioral Engineers. They took their efficiency ratings very seriously and were sometimes upset when they received a low rating. Although regular checks of the BE's methods of teaching are probably valuable, providing trophies for the teams may actually be detrimental because the emotional responses elicited by them interfered with changing the students' behavior.

During the last month of the program, some of the BE indicated they were becoming bored with the daily tutoring routine. To counteract this, BE were given Wednesday as a *free day*. They worked four days a week and were allowed to leave their regular classes on the fifth day and come to the reading room to talk or play games. Another alternative might be to allow BE to work for 6- or 12-week periods with an option to continue if they wish. Boredom, however, was not a problem with the children taught. They eagerly attended the sessions and spontaneously expressed their enjoyment of the training. One of the consultants watched as a little girl, some weeks after the program ended, knocked on the counselor's door to ask, "Mrs. Morris, why don't we get to go to reading any more?"

The fact that the BR program was carried out in a regular public school produced further problems. The path to implementing something that does not fit the *normal* pattern of education is often strewn with so many administrative hurdles that many innovations never make it through the maze. Some school administrators are very suspicious of *outsiders*. This suspicion, and sometimes outright hostility, is often well justified on the basis of the school systems' previous experience. In fact, schools seem to be continually faced with a variety of pressure groups who advocate changes ranging from sensitivity training for first graders to teaching ghetto children judo and birth control techniques. The School Consultation Project has attempted to overcome the inherent and often justified resistance to innovation in the public schools by using a crude shaping process. First, professionals in the target schools identify the target behavior. In the study reported here the school counselor was concerned about the large number of students who read poorly. Second, SCP helps the school develop and implement a small pilot project. In this case ten children were tutored in a corner of the school auditorium. SCP supplied all the material and loaned the school a paraprofessional to manage the program and collect data. The third step, if objective data show the pilot project is effective, is to expand and continue the program with the school system accepting more and more responsibility for a successful program. Here, the school system was impressed with the pilot project data and supplied a room, the reading material, a paraprofessional, and a great deal of positive reinforcement from school administrators. School administrators are now considering expanding the model to all schools in the system during the coming year. However, it should be noted that school administrators responded primarily to the crude pre and posttesting data. The precise daily data had little impact. Besides this function the standardized tests were of little use compared to the crucial daily data.

In summary, the behavioral reading technique described here is simple, inexpensive, and most important -- *effective*. The study joins several hundred other reports of tutoring in the literature. Many have been described in the book, *Children Teach Children*, by Gartner, Kohler and Riessman (1971). The bulk of the reports on tutoring collected little data and used relatively unstructured tutoring techniques, some based on the assumption that the "natural" warmth of the tutor is the most effective approach. However, the log of one student tutor indicates tutors who are not given specific tutoring techniques may develop their own:

> "OCTOBER 25: Kim still reads too fast.
>
> OCTOBER 29: Today I devised a way to slow Kim down. If he reads aloud too fast, I give him a red strip of paper for a speeding ticket. If he reads with expression, I give him a green strip. We are going to make a graph of the strips and try to have it all green.
>
> NOVEMBER 25: Kim does better all the time. I swear he is smarter than me."
> (Fleming, 1969)

EFFICIENCY CHECK LIST

TUTOR: **TEAM:**

DATE:

				1 Reading materials present at beginning.
				2 Tokens present.
				3 Praises often when correct.
				4 Shows enthusiasm.
				5 Gives token after each correct response.
				6 Covers comprehension questions adequately.
				7 Corrects errors.
				8 Corrects errors in *therapeutic* manner.
				9 Transfers reading to other student when error is made.
				10 Gives red chip for error.
				11 Counts chips at end of session and records data.
				12 Counts with students.
				13 Praises and encourages at counts.
				14 Replaces materials in proper place.
				15 Allows student to begin with sentence in which the other student made error.
				16 Helps student complete sentence with errors before transferring reading.
				17 Is on time.
				18 Does tutoring within acceptable noise limits.
				19 Does not quit before end of period.
				20 Keep accurate records (dates & materials).

Each item = 5 *points; Perfect score =* 100 *points*

REFERENCES

Fleming, J. Carl. Pupil tutors and tutees learn together. *Today's Education*, 1969, 58, 48. In Gartner, A., Kohler, M. and Riessman, F., (Eds.), *Children Teach Children*. New York: Harper and Row, 1971.

Gartner, A., Kohler, M. and Riessman, F. *Children teach children*. New York: Harper and Row, 1971.

Gray, B., Baker, R. and Stancyk, S. Performance determined instruction training in remedial reading. *Journal of Applied Behavior Analysis*, 1969, 2, 255-263.

Hayes, E. Prescriptive teaching as a supplement to behavior modification in the remediation of learning disorders. (Doctoral dissertation, University of Southern California.) Ann Arbor, Michigan: University Microfilms, 1967, No. 68-5865.

Libaw, F., Berres, F. and Coleman, J. Evaluating the treatment of learning difficulties. In Long, N. J., Morse, W. C. and Newman, R., (Eds.), *Conflict in the Classroom*. Belmont, California: Wadsworth Publishing, 1965, 505-508.

Silver, A., Hagin, R. and Hersh, M. Reading disability: teaching through stimulation of deficit perceptual areas. *American Journal of Orthopsychiatry*, 1967, 37, 744-752.

Staats, A., Minke, K. and Butts, P. A token reinforcement remedial reading program administered by black therapy-technicians to problem black children. *Behavior Therapy*, 1970, 1, 331-363.

Tomatis, A. *Dyslexia*. (Translated by Agatha Sidlauskas.) Ottawa, Ontario: University of Ottawa Press, 1969.

Willis, J., Morris, B. and Crowder, J. A remedial reading technique for disabled readers that employs students as behavioral engineers. *Psychology in the Schools*, 1972, 67-70.

EFFECTS OF PEER TUTORING ON THE SPELLING PERFORMANCE OF ELEMENTARY CLASSROOM STUDENTS[1]

v. william harris; james a. sherman
david g. henderson; melissa s. harris

The effects of an unstructured peer tutoring procedure on the spelling behavior of elementary school children were investigated. Weekly test gains from pre-test to post-test were consistently higher for tutored word lists than for comparison word lists. Control conditions suggested that interaction between students during the tutorial period was responsible, in part, for the improved spelling performance. Later retests of tutored and comparison word lists showed that the effects of tutoring were still evident after several weeks, although at a reduced level. Systematic replications showed the tutoring procedure to be effective in several different elementary grades.

INTRODUCTION

The use of students to tutor other students in academic subjects is being increasingly used and advocated (Fleming, 1969; Lippitt, 1969; Rossi, 1969; Thelen, 1969; Architeque, 1970), particularly in situations where high student-to-teacher ratios limit the amount of individual instruction the teacher can provide. Several published studies (Horst, 1931; Caditz, 1963; Ellson, Barber, Engle and Kampwerth, 1965; Hassinger and Via, 1969; Taylor, 1969; Landrum and Martin, 1970), as well as several papers read at this conference have indicated that the use of students to tutor other students may result in improved academic performance by these students. We attempted to add to our knowledge about student tutoring by evaluating the effects of an unstructured peer-tutoring system upon the students' performance on spelling exams and to compare the effects of peer-tutoring to the effects obtained when students studied individually.

---

[1] This research was supported by Grant OEG-0-8-522422-4433 from the United States Office of Education to the University of Kansas Support and Development Center for Follow Through. We thank Dr. Don Bushell, Jr. for his support and assistance and Harold Siegrist, principal of Woodlawn School, Lawrence, Kansas, and the entire staff at Woodlawn School for their cooperation in this investigation.

METHOD

Subjects

The majority of the experimental conditions were conducted in a fifth grade classroom composed of twenty students. Prior to the initiation of this study the class had scored a mean of 3.2 years on the spelling component of the Iowa Test of Basic Skills. According to the teacher, a second-year intermediate instructor, the majority of the class had consistently scored less than 75% correct on spelling exams given after completion of the spelling lesson programmed for the week.

Systematic replications of one of the conditions performed in the fifth grade classroom (the tutoring conditions) were conducted in four additional classrooms. A combination second/third grade classroom (24 students), a combination fourth/fifth grade classroom (25 students), and two sixth grade classrooms (22 and 21 students respectively) were selected. Mean class scores on the spelling component of the Iowa Test of Basic Skills, administered at the outset of the year, for each classroom are listed in Table 1.

TABLE I

SCORES ON SPELLING COMPONENT OF THE IOWA BASIC SKILLS ADMINISTERED AT THE OUTSET OF THE YEAR FOR THE VARIOUS CLASSROOMS.

GRADE LEVEL	SCORE = GRADE LEVEL EQUIVALENT
1. *Combination 2/3*	*No Test / 3.0*
2. *Combination 4/5*	*5.1 / 4.7*
3. *6_a*	*5.1*
4. *6_b*	*5.2*

Procedure

The effects of a tutoring system and an individual study system on spelling behavior were investigated in a fifth grade classroom At the outset of each week (Monday) students were administered two 20-word pre-tests, one list after another. Word lists were drawn from BOOK D of the Botel Spelling Series. During the days intervening between pre- and post-tests, students worked on assignments related to both 20-word exams. These assignments were drawn from BOOK D of the Botel Spelling Series. At the end of each week (generally Friday) students were again administered the same two 20-word exams (the post-test) that were given on Monday of that week. Spelling word lists were changed every week.

During the first three weeks in the fifth grade classroom the effects of a tutoring system on spelling performance were observed. Ten minutes prior to administering one of the post-tests, students were given the list of words which comprised one of the 20-word tests, and were allowed to arrange themselves in small groups to tutor each other. The teacher simply told the students to arrange themselves in small groups of from two to four students and to help each other learn the words. The teacher gave no additional instructions as to how the students were to help each other learn the words. On the basis of observation of the tutoring groups, a variety of things occurred. Occasionally one student within a group would serve as spelling examiner for the entire 10-minute tutoring period. More frequently, however, each child within the group would participate as spelling examiner during the tutoring session. Typically, students would write the spelling words as the examiner (tutor) called out the words from the spelling lists. Then, within each group, students would compare answers and make appropriate corrections. Sometimes, particularly when the group was composed of only two students, the examinee (tutee) would vocally respond to words presented by the examiner (tutor).

On the post-test, the tutored list of twenty words was sometimes tested first and at other times second. Whether or not the tutored word list was post-tested first or second, the 10-minute tutoring period always immediately preceded the test on the tutored list of twenty words.

During the three weeks in which tutoring was employed, spelling performance on the tutored list of twenty words was compared to spelling performance on the list of twenty words which was not tutored. The list chosen for tutoring was determined by a coin flip.

To evaluate whether the effects of tutoring were due to interactions between students during the 10-minute tutoring sessions prior to the post-test, or simply the opportunity that students had to study one list for ten minutes prior to the post-test, an independent study condition was arranged for three weeks. In the independent study condition, each student was given a list of words which comprised one of 20-word tests ten minutes before the post-test on that list of words. The students were told to study this list of words individually. All other procedures remained the same as in the tutoring conditions. The list of words chosen for independent study was determined by a coin flip.

Following three weeks of the independent study condition, a series of one-week conditions were employed as follows: *tutoring*, *independent study*, *tutoring*, and *independent study*.

Tutored and comparison word lists and independently studied and comparison word lists were retested from two to five weeks following administration of the post-test in the fifth grade classroom. This was done to evaluate the possible effects of the tutoring and independent study procedures over a somewhat longer period of time.

To evaluate whether the effects of the tutoring procedure were replicable across different classrooms, the same tutoring procedures employed in the fifth grade classroom were used in a combination second/third grade classroom, a combination fourth/fifth grade classroom and two sixth grade classrooms. These procedures were in effect for one week in each classroom. As in the fifth grade, two 20-word spelling tests were pre- and post-tested for three of the classrooms during this week. In one of the classrooms (the combination second/third grade), two 10-word spelling tests were used.

Each student's spelling pre- and post-test was graded and returned on the same day that the test was administered. A percentage correct was determined for each student on each word list as well as a mean percentage correct for the entire classroom on each word list. The mean percentage correct for each classroom was determined by dividing the total number of correct words spelled for all students in the classroom by the total number of words possible for all students in the classroom. No attempt to spell a word was scored as an incorrectly spelled word.

Pre- and post-test grading reliability was assessed at least once during each of the conditions in the fifth grade classroom. A pre- and post-test grading reliability check was completed for each of the other four classrooms during the only week in which spelling performance was observed. The degree to which two independent graders agreed on whether a word was correctly or incorrectly spelled was determined by dividing the total number of words (for all students) agreed upon by the total number of words possible (for all students). Reliability scores for the pre- and post-tests on both the experimental and comparison word lists averaged 99% across all of the conditions conducted in the fifth grade classroom. Reliability scores for the pre- and post-tests on both word lists also averaged 99% in the other four classrooms.

RESULTS

Mean classroom percentage scores were calculated over all students that took both the pre-test and post-test during any given week. Weekly classroom gains from pre-test to post-test for each set of twenty words are displayed in the upper graph of Figure 1. Weekly test gains of the tutored or individually studied lists of words are represented by the solid line. The comparison list of words for each week is represented by the dotted line. Bars in the lower graph of Figure 1 represent the classroom test gains of the tutored or independently studied word list over the comparison set of words tested for that week.

During the first three weeks of the study the tutoring procedure was used for one set of twenty words. The post-test gain for the tutored set of words ranged from 14% to 25% greater than the post-test gain on the non-tutored set of words over the three weeks. During the next three weeks one list of words was studied independently for the same

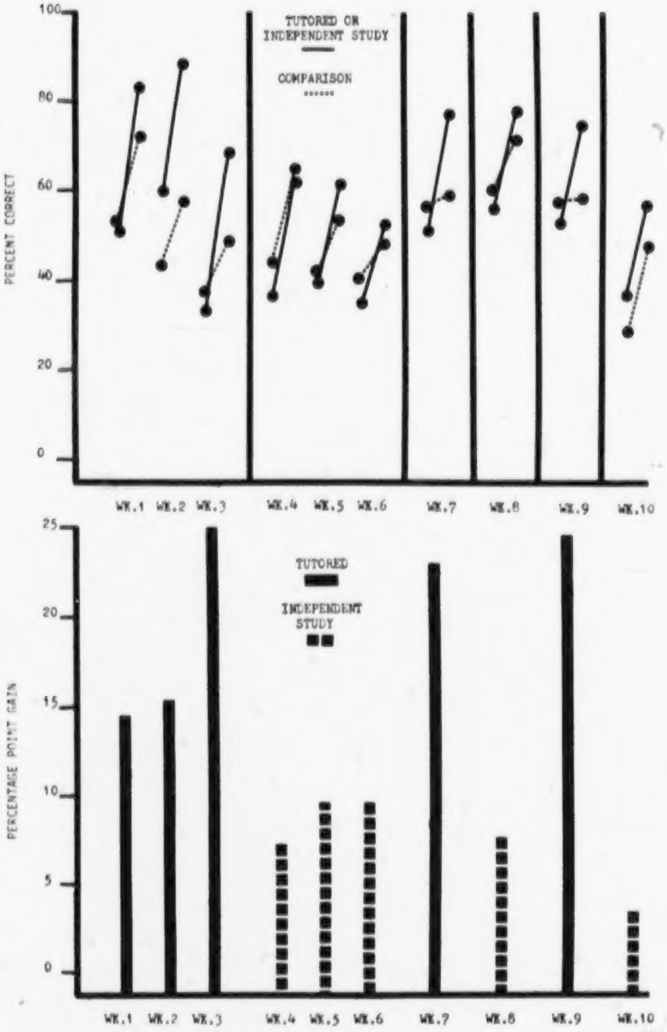


Fig. 1. **UPPER SET OF AXES:** mean percent correct for all students on each weekly pre-test and post-test for tutored, independently studied, and comparison word lists. **LOWER SET OF AXES:** mean percentage point gain for tutored or independently studied word list over the comparison word list during each week.

amount of time as one list had been tutored in the first three weeks. The independently studied list of words gained from 7% to 9% over the comparison lists of words on the post-test over these three weeks. When the independent study procedure alternated with the tutoring procedure each week, the post-test gains for independently studied words were 8% and 3% larger as versus the comparison words, while the gains for tutored words were 23% and 24% higher than for the comparison words.

Figure 2 shows the percentage of students scoring within ten percentage-correct intervals on the post-tests over all tutored and independently studied word lists. As shown in this figure, more students scored higher on the tutored word lists than on the individually studied word lists, although there were also slight differences in the initial pre-test scores for tutored words as compared to the individually studied words. The mean accuracy over all pre-tests for tutored words was 44% and was 40% for individually studied words.

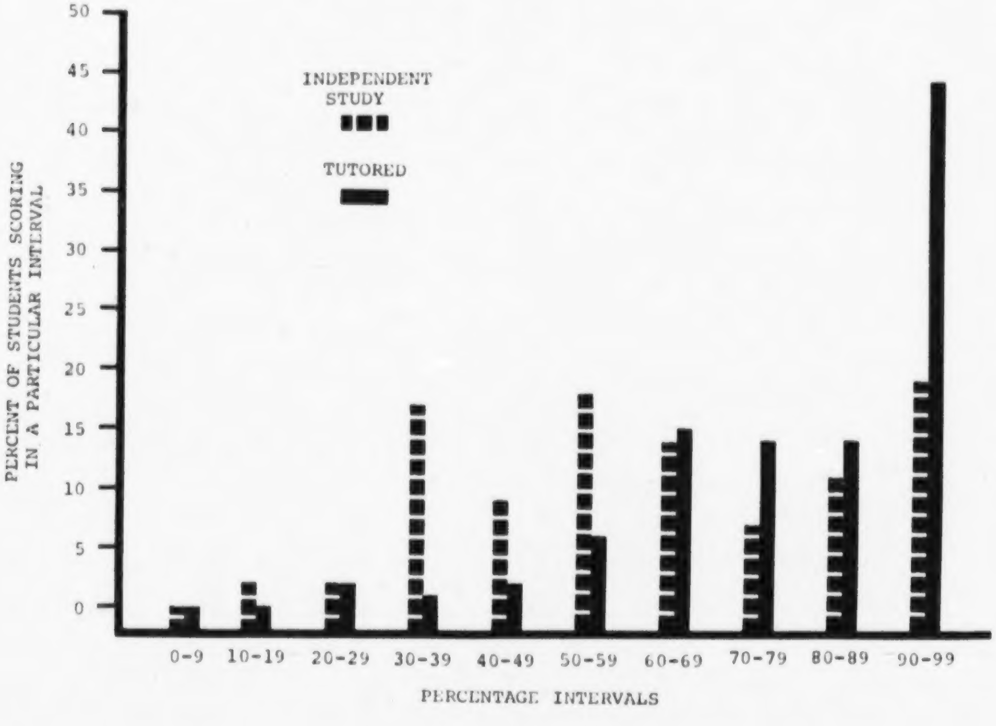


Fig. 2. Percentage of students scoring within ten percentage-correct intervals on the post-tests for all tutored and independently studied word lists.

A tutored or independently studied word list and the appropriate comparison word list were retested from two to five weeks following administration of the post-test procedures. This retest procedure provided data for determining the effects of the tutorial and individual study procedure on retention. These results are shown in Table 2.

TABLE 2

PERCENTAGE CORRECT ON TUTORED, INDIVIDUALLY STUDIED AND COMPARISON WORD LISTS WHEN RETESTED.

WEEK OF PRE-POST TEST	CONDITION	MEAN CLASSROOM PERCENTAGE			WEEK RETEST
		PRE-TEST	POST-TEST	RETEST	
③	*Tutored List*	33	68	49	8
	Comparison List	38	48	41	
⑦	*Tutored List*	52	77	66	11
	Comparison List	57	59	59	
⑤	*Individual Study List*	40	61	49	7
	Comparison List	41	53	54	
⑥	*Individual Study List*	35	52	44	11
	Comparison List	41	49	42	

Both tutored sets of words retested maintained higher mean percentage scores than their respective retested comparison word list. However, decreases in mean classroom scores from post-test to retest were greater for tutored words than for comparison word lists. One list of retested words that had been independently studied maintained a slightly higher mean accuracy than the comparison word list. The other retested list of individually studied words dropped to an accuracy lower than the comparison word list. Like the tutored word sets, individually studied words showed greater decline from post-test to retest than did comparison word sets.

The procedures of the tutoring system used in the fifth grade classroom were replicated in four separate classrooms. Figure 3 shows the pre- to post-test gains for each of the four classrooms selected for study. Solid lines indicate the gains made on the tutored list of words. Dotted lines indicate the gains made by the comparison list of words. The bars at the base of each graph display the mean percent pre- to post-test gain of the tutored list of words over the comparison list of words. In each of the four classrooms tutored words showed a larger gain from pre-test to post-test than the comparison word lists. Tutored words in

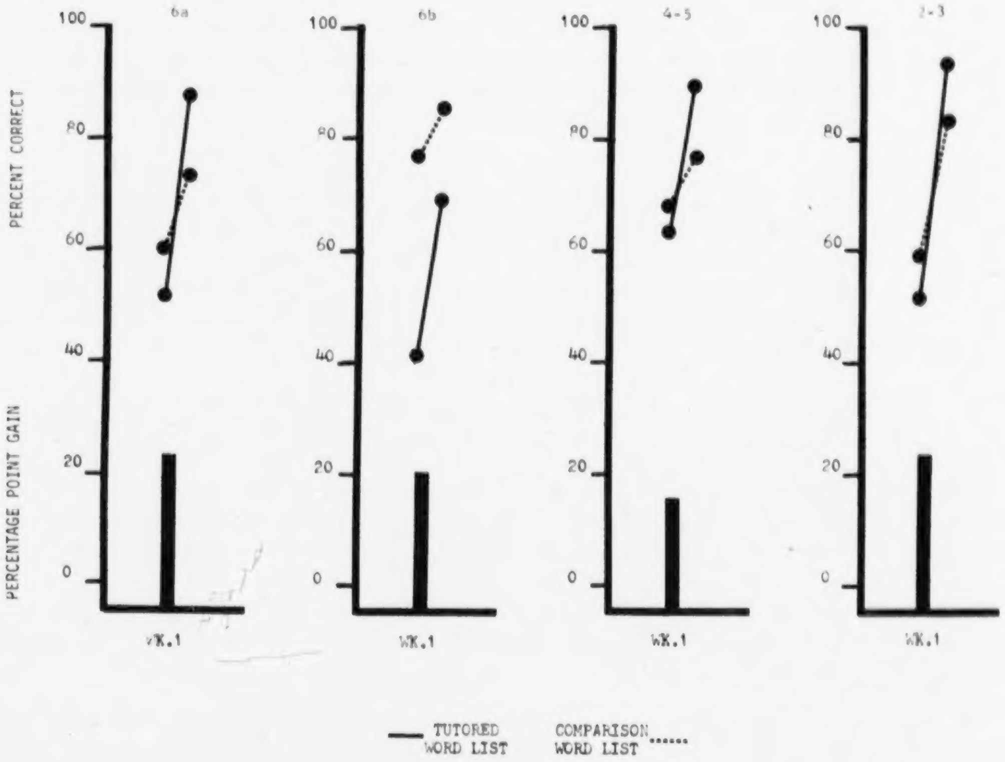


Fig. 3. Mean percent correct for all students in each of four classrooms on the pre-test and post-test for tutored and comparison word lists. The bar at the base of each set of axes represents the mean percentage point gain for the tutored word list over the comparison word list.

the combination second/third grade classroom showed a 23% gain over the comparison word list. In the combination fourth/fifth grade classroom tutored words scored a 16% gain over the comparison word list. In classrooms 6_a and 6_b tutored words scored gains of 23% and 20% respectively.

DISCUSSION

The present investigation evaluated the effects of a tutoring procedure on the spelling behavior of elementary school children. Weekly test gains from pre-test to post-test were consistently higher for tutored word lists than for comparison word lists. Independent study control conditions indicated that the effect of the tutoring procedure could not be attributed solely to the opportunity that students had to study one list for ten minutes prior to the administration of the post-test. Retests of tutored word lists and their respective comparison word lists showed the tutoring effects to be maintained over a four to five week period, although at a reduced level. Systematic replications of the tutorial procedure in several additional classrooms indicated that the tutorial procedure was effective across a large number of students in several different elementary grades.

Because the tutorial procedure was relatively unstructured (i.e., the teacher simply asked the students to help each other learn the words) it was impossible to distinguish between the differential effects the tutorial procedure may have had on the tutor or tutee per se. Other investigators, however, have reported that a tutorial procedure may be effective in producing academic gains for both the tutor and the tutee (Landrum and Martin, 1970; Hassinger and Via, 1969).

One characteristic of the tutorial procedure potentially responsible for the behavioral effect was the feedback given by students to other students regarding the correctness of a spelled word. Immediate feedback occurred particularly when a tutorial group was composed of only two students where the tutee would vocally respond to words presented by the tutor. Less immediate feedback occurred when the students wrote the spelling words as they were called out from the spelling list. A second characteristic of the tutorial procedure potentially responsible for the effect was writing or spelling out loud the words to be tested prior to the test situation.

An important characteristic of the present tutoring procedure was that it required little time and no additional materials to implement. All of the teachers who used the procedures found them both practical and convenient and advocated their continued use. The fact that the tutoring procedures seemed to work in the absence of specific training procedures for tutors, suggests that elementary school students already have some tutoring skills, at least in areas such as spelling where drill and repetition may be useful teaching techniques. Similarly unstructured tutoring procedures appeared to be effective, however, in other subject areas employing older student tutors (Horst, 1931; Caditz, 1963; Hassinger and Via, 1969; Landrum and Martin, 1970; Taylor, 1969).

REFERENCES

Archibeque, Joe D. Utilizing the advanced Spanish student as a classroom tutor. *Hispania,* March, 1970, 53, 70-72.

Caditz, Rebecca. Using student tutors in high school mathematics. *Chicago School Journal,* 1963, 44, 323-325.

Ellson, D. G., Barber, L., Engle, T. L. and Kampwerth, L. Programmed tutoring: a teaching aid and a research tool. *Reading Research Quarterly,* 1965, 1, 77-127.

Fleming, J. Carl. Pupil tutors and tutees learn. *Today's Education,* October, 1969, 58, 22-24.

Hassinger, Jack and Via, Murray. How much does a tutor learn through teaching reading. *Journal of Secondary Education*, January, 1969, 44, 42-44.

Horst, H. M. History of student tutoring at West High School, Akron, Ohio. *Junior-Senior High School Clearing House,* 1931, 24, 69-81.

Landrum, John W. and Martin, Mary D. When students teach others. One-to-one project, Los Angeles Co. Schools. *Educational Leadership,* February, 1970, 27, 446-448.

Lippitt, Peggy. Children can teach other children. *The Instructor,* May, 1969, 78, 41, 99.

Rossi, Timothy P. Help: students teach students. *Reading Improvement,* Fall, 1969, 6, 47-49.

Taylor, Ronald G. Tutorial services and academic success. *The Journal of Educational Research*, January, 1969, 62, 195-197.

Thelen, Herbert A. Tutoring by students. *The School Review*, September, 1969, 77, 229-244.

* * *

EFFECTS OF HAVING ONE REMEDIAL STUDENT TUTOR ANOTHER REMEDIAL STUDENT[1]

michael davis

Because of increasing demands on the professional teacher and because educators are becoming more responsive to student needs for individual tutoring, paraprofessional teacher aides, parents, and college students have been used increasingly to supplement the public school teaching staff. Another source of manpower for the teaching staff is other students who have already mastered the material to be taught or who have been especially trained and equipped as contingency managers (Surratt, Ulrich, and Hawkins, 1969). Other procedures for using students as teachers simply employ classroom contingencies that result in peer tutoring (Conlon, Hall, and Hanley, 1972; Hamblin, Hathaway, and Wodarski, 1971). In some cases special training has been provided for students to manage the contingencies of their peers (Bailey, Piotrowski, and Johnson, 1972; Harris, Sherman, and Henderson, 1972; Willis and Crowder, 1972) or both peers and teachers (Graubard, Rosenberg, and Miller, 1971). At this point it is interesting to note that all these references are to papers from either the 1971 or 1972 Conference on Behavior Analysis in Education with the exception of the Surratt study.

The use of one student to teach another student would be a particularly appealing procedure if the teaching-student (the tutor) were to show some improvement in the quality of his academic behaviors as a result of assisting others. The present project employed a remedial student as a tutor of other remedial students and observed changes in the quality and quantity of the tutor's academic performances and changes in the quality of the performances of his students.

METHOD

Subjects

The two tutors or contingency managers were sixth grade boys who scored at least two grade levels low on the *Stanford Achievement Test* (*SAT*) and

---

[1] The author wishes to thank Mr. K. Schlappi, Mrs. G. Price, and Mrs. L. Powell of the Taylorsville Elementary School for their cooperation. Mr. John Anderson and Mr. Ross VanVranken who were raters greatly assisted in the data reduction tasks and operation of the program.

were judged by their teacher to be remedial students. The *SAT* scores (reading scales) for these two boys and their ages are shown in Table I.

TABLE I

MANAGERS' AND STUDENTS' AGES (IN PARENTHESES)
AND STANFORD ACHIEVEMENT TEST SCORES (SAT)

Managers	*SAT*	*Students*	*SAT*
Kevin (12)	3.6	Edith (8)	.6
		Mark (9)	1.8
Curtis (13)	2.2	Johnny (9)	1.2
		Rick (9)	1.8

The *SAT* scores were grade-level scores relative to a grade-level norm. The student managers were each responsible for two third grade students who were also judged to be remedial reading students. The ages and *SAT* scores for these students are also shown in Table I.

Dependent Variables

The measure of academic performance was the number of answers to questions and exercises provided in the Science Research Associates Reading Laboratory Power Builder (SRA) series of programmed materials (Parker and Scannel, 1961). Two summary indices were calculated from the number of answers. First, the *percent of items completed* on any school day was calculated by dividing the total number of item answers by the total number of possible answers. Second, the *percent correct answers* was calculated daily by dividing the total number of correct answers by the total number of answers that were possible. The percent completed index indicated the quantity of performance, and the percent correct index indicated the quality of performance. Written responses (answers) were required of the tutors at all phases of the project. Spoken or written responses were required of the students depending upon the phase of the project.

Reliability

The students' spoken answers and the tutors' praise responses (consequences) were tape recorded during the students' tutoring sessions for subsequent independent rating by two raters. Indices of agreement were calculated by dividing the number of rater agreements of student answers or tutor consequence responses by the total number of possible SRA answers or total number possible, appropriate consequences. These indices of agreement could range from 0.00 to 1.00 with 1.00 indicating perfect agreement. The indices were calculated daily and ranged from .82 to .98 with a median index equal to .89. Written answers to SRA exercises were graded by the students' or the tutors' teacher or by the tutors and were checked again by one of the raters. Three discrepancies between a rater and a teacher occurred out of over five thousand answers, and eight discrepancies between a manager and a rater occurred out of over four hundred answers.

Manager Training and Tutoring Contingency

The managers were trained to tutor by having the raters and the author model and tutors engage in role playing correct and incorrect student responses and appropriate and inappropriate praise consequences and prompting responses. The praise responses were taken from lists of *spoken approval responses* as described by Madsen and Madsen (1970, pp. 116-118). Prompting responses generally involved providing the student with a spoken cue that had phonic similarity or a similar meaning to the correct answer. Answers provided by the student following a prompt by the manager were not counted by the raters.

The managers were told that their teaching would ~~be~~ evaluated by the raters who would listen to the tape recordings. The managers were evaluated on the basis of how well they:

(1) identified the correct answers;

(2) praised or token-reinforced those answers;

(3) provided appropriate prompts when the student erred;

(4) escorted the third grade students to and from the tutoring room; and

(5) set up the tutoring room for use.

If the tutors completed all their assignments they were awarded a maximum of 25 points per day. Failure to reinforce correct answers, put chairs away, and so forth, resulted in a reduction of that amount of points. The points earned by the tutors were accumulated until the end of the program. The tutors were told that they could arrange to exchange their points for a field trip or some comparable activity at the end of the program.

The Tutoring Program

The tutoring program consisted of daily one-to-one contingency management sessions where the students were required to provide spoken answers to SRA exercises. These sessions were held in a storage room where teachers' materials were stored and where teachers prepared visual aids and curriculum materials. The tutor sat facing the student who was seated in a tablet armchair. To the right of the student was a 12 inch square vertical wooden panel on which a microphone and a plastic token receptacle were mounted. The two tutors worked simultaneously, and each tutoring session was recorded on a separate channel of a two channel tape recorder. The tutor-student pairs were separated by a four foot high storage cabinet so that visual contact between pairs was impossible and auditory noise was minimal. When the student completed all exercises in the assigned SRA booklet the tutoring session was terminated, and the student returned to his classroom.

The program was conducted as a series of manipulations in contingencies that affected the behavior of the student, the tutor, or both. The student contingencies had to do with the scheduling of praise and token consequences and the setting in which the student completed the SRA exercises. The tutor contingencies had to do with scheduling of events that were potential reinforcers or accelerators of academic performances including free time from the classroom and practice at the SRA materials. The sequence of phases of the program has been briefly summarized in Table 2, and the procedures in each type of manipulation are described below. Briefly there were four types of procedures involved in the project.

1▶ BASELINE - The tutors were instructed to make no response or a neutral response to the student's answers. Neutral responses consisted of asking the student to answer the next item in the exercises. At the end of the session the tutor announced, "Okay, it's time to go back to your room." The tutoring contingencies were in effect for the tutors.

2▶ REINFORCEMENT - During all reinforcement phases the tutor consequated the student's correct responses on a compound schedule where every correct response was praised and every fifth correct response was reinforced with a token. That is, a compound continuous praise, fixed ratio five token schedule (comp CRF FR 5) was in effect. The tokens were exchangeable for nonmaterial backup reinforcers in a token economy in the third grade classroom. All other procedures were the same as during baseline, and the tutoring contingencies were in effect for the tutors.

3▶ EXTINCTION - Extinction procedures were a return to baseline conditions of praise consequation. However, either of two conditions of token reinforcement were employed: the FR 5 was continued (as in *Extinction II* in Table 2), or token reinforcement was discontinued (i.e., token extinction as in *Extinction I* in Table 2). Under the token extinction condition the tutor told the student, "Today I can't give you any tokens for correct answers." The tutoring contingencies were in effect for the tutors.

4▶ NO PROGRAM - No program procedures involved instructing the tutors and the students that the third grade teacher wanted to watch the students work for a while. The students provided written answers to SRA exercises in the classroom, and no praise consequences were scheduled. Either of the two conditions of token reinforcement was employed, FR 5 or extinction. During no program phases the tutoring contingency was altered in one of several ways as follows:

- 1● *No Program I* - The tutors were told that the teacher wanted to watch the students work and that the program was going to be discontinued for a while. The tutors continued to be assigned SRA exercises in their own classroom during all *No Program* phases but they were not allowed to leave the classroom.

- 2● *No Program II* - The tutors were told that the teacher wanted to watch the students work, but that the raters could use their help in the third grade classroom. The tutors were given a contrived task that involved transcribing numbers from one type of form to another form. This procedure allowed the tutors to have free time from their classroom contingent upon completion of their SRA exercises.

- 3● *No Program III* - The program was discontinued, the tutors were told that they could again help the raters, and the task they helped with was scoring the third grade students' SRA exercise answers. This allowed the managers to have practice with the SRA materials in addition to contingent free time.

TABLE 2: *SEQUENCE OF PROGRAM PHASES and CONTINGENCIES*

PHASE OF PROGRAM	STUDENTS' CONTINGENCIES		MANAGERS' CONTINGENCIES
	Praise	*Tokens*	
1. BASELINE	ext*	ext	Tutoring Contingency (TC)
2. REINFORCEMENT I	crf**	FR 5***	TC
3. EXTINCTION I	ext	ext	TC
4. REINFORCEMENT II	crf	FR 5	TC
5. EXTINCTION II	ext	FR 5	TC
6. NO PROGRAM I	ext	FR 5	Returned to Classroom
7. REINFORCEMENT III	crf	FR 5	TC
8. NO PROGRAM II	ext	ext	Free Time from Classroom
9. EXTINCTION III	ext	ext	TC
10. NO PROGRAM III	ext	ext	Phase 8 Contingencies *plus* SRA Practice
11. REINFORCEMENT IV	crf	FR 5	TC
12. NO PROGRAM IV	ext	FR 5	Phase 10 Contingencies *plus* Points
13. REINFORCEMENT V	crf	FR 5	TC

*ext - extinction procedure
**crf - continuous reinforcement procedure
***FR 5 - reinforcement of every fifth correct response

RESULTS

Student Performance

The effectiveness of the program as a teaching procedure was evaluated in terms of the quality of performance by the students. These data are shown in Figures 1 and 2 where the percent correct responses have been plotted for each daily session. Also, shown are the percentage of correct responses that were praised by the tutors (triangles). The praise data shown are average percentages for blocks of five sessions. The percentage of correct responses that were consequated with praise during reinforcement phases and during extinction phases were very similar to the definitions of continuous reinforcement (100%) and extinction (0%).

The first series of experimental manipulations shown from left to right on Figures 1 and 2 was an *A - B - A - B* sequence where *Baseline* and *Extinction I* were the *A* conditions and *Reinforcement I* and *II* phases were the *B* conditions. All students except Mark improved above baseline to near perfect (100%) performance during *Reinforcement I*, and Mark's performance was considerably higher than during baseline. Removal of the token and praise consequences during *Extinction I* resulted in a decrement in the quality of performance of all students and reinstating the program contingencies during *Reinforcement II* resulted in recovery of previous reinforcement phase quality of performance. That is, the program contingencies in effect during reinforcement phases consistently resulted in high quality performance as compared to baseline or extinction phases.

The next series of manipulations was instated to evaluate several variations of the tutoring program involving either the removal of the praise consequences or removal of the praise consequences and changes in the setting in which students completed their SRA exercises. This sequence was the *Reinforcement II* - *Extinction II* (FR 5) - *No Program* (FR 5) - *Reinforcement III* series in which the reinforcement phases were the comparison conditions. For all students the removal of the praise consequences during *Extinction II* resulted in a decrease in the quality of performance relative to the reinforcement phases. Instituting the *No Program* phase involved having the students write answers to the SRA exercises in the third grade classroom without having the tutor present. In addition, there was approximately a 20-minute delay in the delivery of tokens. The quality of performance of all students except Edith's fell below the preceding extinction phase levels and the reinforcement phase levels. Edith's performance remained below the level of quality obtained during the reinforcement phases. These results indicate that removal of praise consequences disrupted the high-quality levels of performance maintained by the program contingencies.

A clearer picture of the effects of changes in the setting was obtained in the next sequence of manipulations where the token reinforcement contingencies were not confounded in the manipulations. This sequence

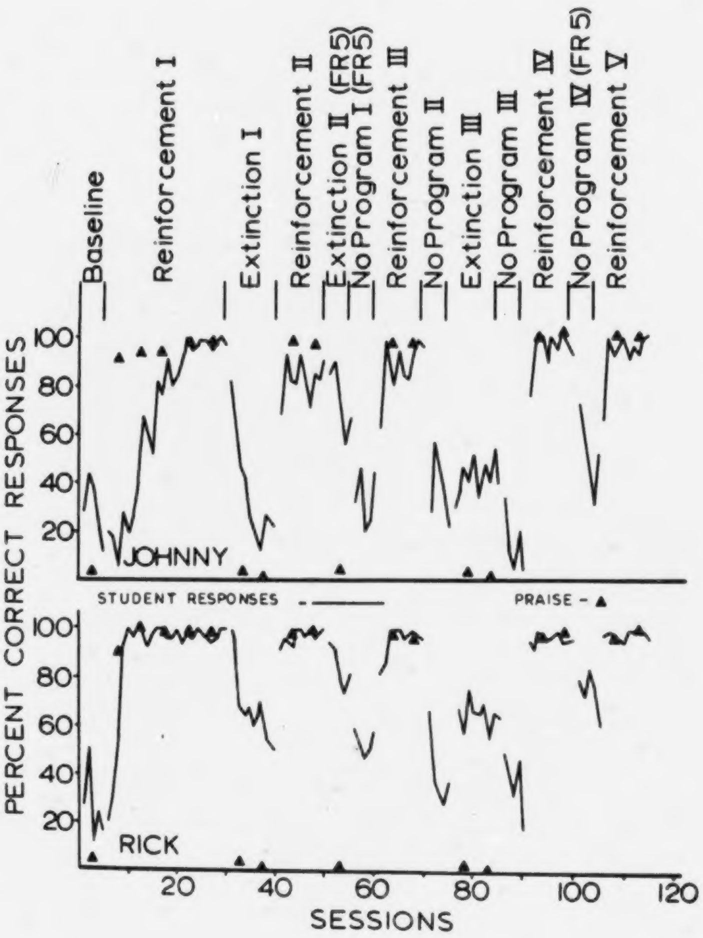


Fig. 1. Percentage correct SRA performances by the third grade students and percent of correct responses that were praised by the tutors. The percentage correct data are shown for each day of the program, and the praise data are average percentages for each block of five sessions. The data shown are from students tutored by Kevin. Phases of the program are noted atop the figure. FR 5 notations accompanying an extinction or no program phase indicate that token reinforcers were scheduled during that phase.

involved the *No Program II - Extinction III - No Program III* phases. Two students, Rick and Mark, answered more items correctly during the extinction phase than during either of the no program phases. Edith and Johnny did not show any consistent improvement during the extinction phase as compared to the *preceding* no program phase. However, the change from extinction (the tutoring program setting) to no program (the classroom setting) resulted in a clear decrement for both of these students. These data indicated that a higher quality of performance occurred in the tutoring program setting than in the classroom setting when neither praise nor token consequences were scheduled.

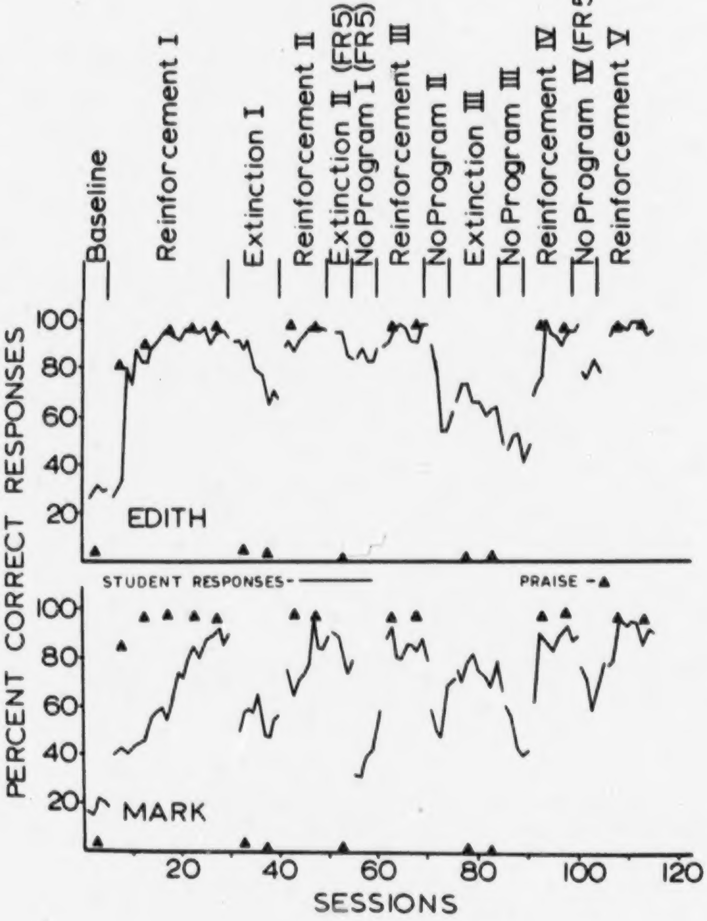


Fig. 2. Percentage correct SRA performances by the third grade students and percent of correct responses that were praised by the tutors. The percentage correct data are average percentages for each block of five sessions. The data shown are from students tutored by Curtis. Phases of the program are noted atop the figure. FR 5 notations accompanying an extinction or no program phase indicate that token reinforcers were scheduled during that phase.

The final sequence of three phases was planned as a comparison of the tutoring program with a token reinforcement contingency that was very similar to that scheduled for the other students in the third grade classroom. This series of manipulations involved the *Reinforcement IV - No Program IV* (FR 5) - *Reinforcement V* phases. All students' quality of performance declined markedly during the token reinforcement contingency (no program phase) as compared to the tutoring program (reinforcement phases).

Tutor Performances

The tutors were required to write answers to each item in the assigned SRA exercises prior to leaving their classroom and beginning the tutoring sessions. This contingency was not in effect prior to the beginning (baseline) phase of the program nor during the first no program phase. The percentage of the assigned exercises that were completed during some of the phases of the experiment has been summarized in Table 3.

TABLE 3

AVERAGE PERCENTAGE OF SRA EXERCISES COMPLETED BY TUTORS

	Phases of the Program					
TUTOR	PRE-BASELINE (5 SESSIONS)	REINFORCEMENT I (SESSIONS 11-30)	EXTINCTION I	NO PROGRAM I	REINFORCEMENT III	ALL SUBSEQUENT PHASES
KEVIN	*67%*	*100%*	*100%*	*84%*	*100%*	*100%*
CURTIS	*52%*	*96%*	*100%*	*72%*	*100%*	*100%*

During the five sessions prior to beginning the program (pre-baseline) the tutors completed 67% or less of their assigned exercises. During *No Program I* they completed less than 85% of the exercises. In contrast, beginning with the eleventh session during the first reinforcement phase Kevin completed 100% of the items when he was at school and when he was required to complete all exercises in order to leave the classroom. Curtis completed 96% of his exercises during the final twenty sessions of the first reinforcement phase and completed all exercises after session 26 when the contingency on completing exercises was in effect. The contingency on completing the SRA exercises appeared to be highly effective in terms of the tutors completing their assigned exercises.

The quality of the tutors' performance is summarized in Figure 3. The percent correct responses each day of the program are shown. The tutors' participation in the program (*Baseline* through *Extinction II* phases) resulted in an increment in the quality of performance up to session 30 for Kevin and to session 42 for Curtis regardless of the type of contingencies that were scheduled for the third grade students (i.e., baseline, reinforcement, extinction). However, discontinuing the program *(No Program I)* resulted in a decrement in both tutors' performances. Returning to the program (*Reinforcement III* phase) resulted in an immediate reappearance of the high quality performance.

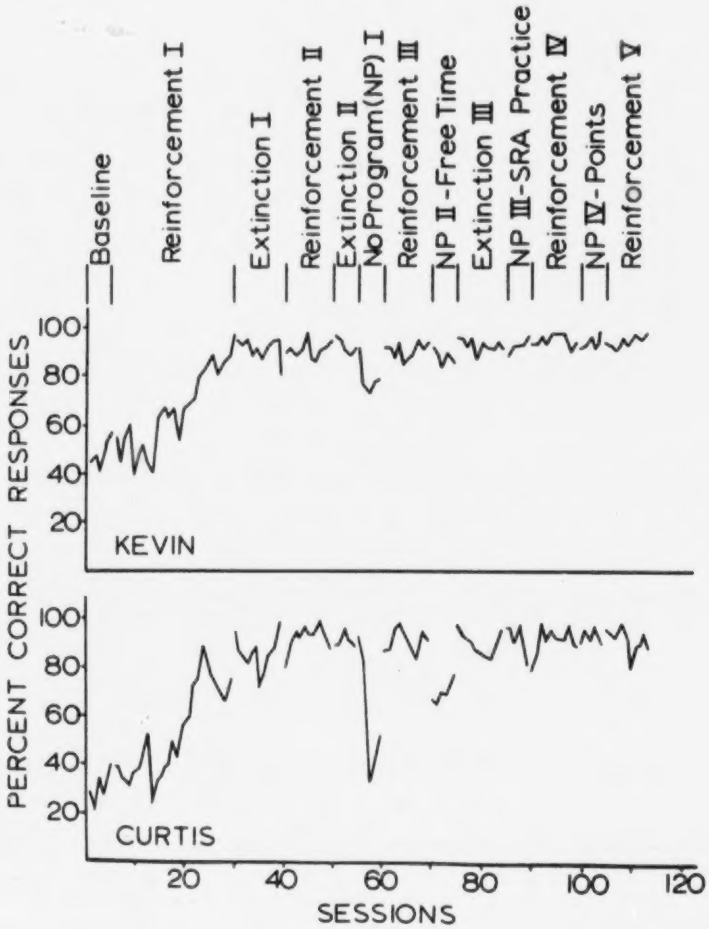


Fig. 3. Percentage of correct SRA performances by the sixth grade students (tutors) for each day of the program. Phases of the program are noted atop the figure.

Several variations in the *No Program* contingencies (NP) were scheduled as an attempt to assess the relationship between these variations in NP and the quality of the tutors' performance. The first NP variation that was scheduled involved making *free time* from the classroom contingent upon the tutors' completion of the SRA assignment (NP II = FREE TIME). Curtis' performance declined in quality during NP II as compared with the *Reinforcement III* phase and the *Extinction III* phase. However, the decline was not as great as observed during NP I, and the very slight decline in Kevin's performance may have been attributable to day-to-day variation rather than the effect of the NP II manipulation. In either case, the disruption occurring during NP II was considerably less than during NP I where no contingency was scheduled.

The second variation that was scheduled involved the free time contingency scheduled in NP II *and* in addition providing the tutors with practice scoring the SRA materials. There was no detectable decrement in the quality of performance for either tutor relative to the preceding extinction and subsequent reinforcement phases. This suggests that for Curtis SRA practice in combination with the free time contingency was adequate to result in quality performance that was equivalent to the quality obtained during program phases.

Absenteeism

Absenteeism was recorded daily by the third grade teacher and the sixth grade teacher as a matter of school policy. These records were used to obtain absenteeism data on the two tutors and four students for the entire school year. From these data the percentage of days when the students or tutors were absent was calculated by dividing the number of absences by the total number of school days. The percent absences are summarized in Table 4.

TABLE 4

PERCENT ABSENCES & TOTAL DAYS
IN EACH TYPE OF PROGRAM PHASE (IN PARENTHESES)

	Type of Program Phase		
STUDENTS	PRE-PROGRAM (*40*)	PROGRAM (*95*)	NO PROGRAM (*20*)
KEVIN	*22.5*	*5.3*	*10*
CURTIS	*25.0*	*3.1*	*10*
EDITH	*12.5*	*6.3*	*5*
MARK	*15.0*	*5.3*	*10*
RICK	*15.0*	*8.4*	*0*
JOHNNY	*37.5*	*16.9*	*5*

There were 40 days in the pre-program phase, 95 days in the program phase (including reinforcement and extinction phases), and 20 days in the *No Program* phase. During the 40 days prior to the beginning of the baseline phase all third grade students had a rate of absenteeism of at least 12.5% or one out of every eight days. One student, Johnny, was absent 37.5% of that time. The tutors had absenteeism rates of about 20% to 25%. Relative to the all-school norm (average) of 5.2% and the school district norm of 6.1% these rates were extremely high. With the exception of Johnny, all students' and tutors' absenteeism rates declined to near the school district norm during the program phases. Johnny's absenteeism was reduced markedly from its pre-program levels.

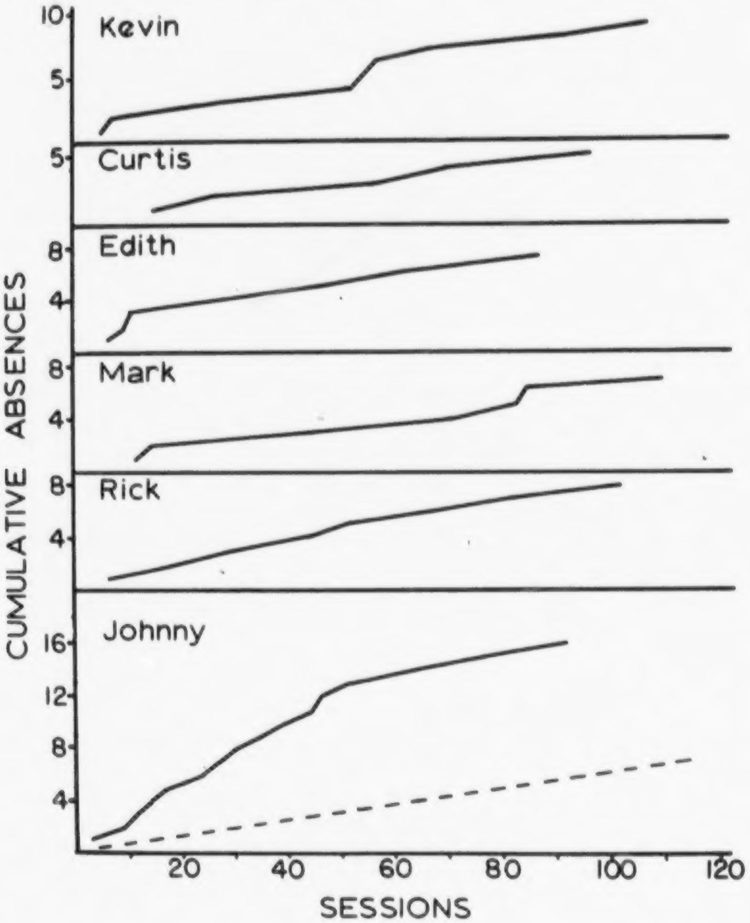


Fig. 4. The cumulative number of absences occurring during program sessions. The top two panels are the tutors' data. The dotted line in the bottom panel indicates the average rate of absenteeism reported by the school district for the same school year.

A finer grain analysis of absenteeism during the program and no program phases can be made from Figure 4. The number of absences is plotted cumulatively, and the dotted line in the bottom panel is the school district norm (average) rate of absenteeism. All rates of absenteeism were similar to the school district norm except Johnny's. Up to about the fiftieth session Johnny's rate was approximately four times the school district norm, and after the fiftieth session his rate was comparable to that norm.

DISCUSSION

There were two criteria of effectiveness of the present tutoring program. First, the program contingencies consistently resulted in improved performance by the third grade students. That is, the combination of praise and token consequences for correct answers resulted in high quality performance by the students as compared to no consequences (baseline and extinction phases in Figures 1 and 2) and to programs that were scheduled in the classroom setting. Secondly, the tutors completed more of the assigned SRA exercises when their participation in the tutoring program was contingent on their completion of exercises than when those contingencies were not in effect (Table 3). In addition, the quality of the tutors' performance on the SRA exercises appeared to be functionally related to their participation as tutors (Figure 2). These results suggest that the program was highly effective as an instructional procedure.

Additional manipulations of program variables were made. First, to determine if the praise consequences contributed to the effectiveness of contingencies effecting the students' performance the praise consequences were discontinued (*Extinction II* in Figures 1 and 2). The removal of the praise resulted in a marked reduction in the quality of all students' performance. Pilot data from a previous variation of the present program indicated that praise alone was not effective in *generating* consistently high quality performance. Discontinuing praise during the *Extinction II* phase was a different procedure than the pilot procedures and was made to determine if the praise was involved in the *maintenance* of the students' high quality performance as contrasted with the effectiveness of praise to condition a similar level of performance. There was the possibility that praise alone was not effective during the pilot program since the third grade students were participating in an extensive token economy in their third grade classroom (Davis, Morris and Price, 1970). That is, perhaps praise consequences were not very effective relative to token consequences that were readily available in the third grade classroom.

A second set of manipulations was made to determine if the effects of the program were the result of simply a change in setting for the third grade students. That is, the students' improved performance could have been maintained, in part, because of the tutoring setting and not the praise or token contingencies. Comparison of data from *Extinction II* and *No Program I* phases suggested that under the FR 5 token schedule three of four students performed at a lower level of quality in the classroom as compared to the tutoring setting. These effects could be attributed to the delay in presentation of token consequences that was necessary in the classroom setting. However, comparison of data from *Extinction III* phase with *No Program II* and *III* phases suggested that the poorer performance in the classroom setting occurred independently of the token consequation or praise. The differences in setting were not simple, single variable differences. First, the tutors in the tutoring situation imposed some demand characteristics on the student to attend to the SRA materials and to respond to the items in those materials at a fairly high rate. In the classroom the student was fairly free to talk to neighboring students and to dawdle. Also, an oral response was required of the student in the tutoring setting, and a written response was required in the classroom. The contribution of these or other potential variables was not determined. Rather, the point is that the tutoring setting alone appeared to maintain some minimal level of improved performance as compared to the classroom setting. However, this effect may be partly a function of the students having experienced many reinforced trials since baseline levels of performance were recovered during the first extinction phase for only one student, Johnny, and baseline levels were not recovered during subsequent extinction (both praise and token) phases for any of the students.

Since a token reinforcement program had been operative for several years in the third grade classroom (Davis, Morris and Price, 1970), the tutoring program was seen as a special remedial program that might have been helpful for some students who were only minimally affected by the classroom token contingencies on quality performance in reading. At the end of the tutoring program, after a large number of reinforced responses, these remedial students may have been able to continue responding under token reinforcement contingencies similar to those in effect for other students in the classroom. To assess generalization to a token economy the final no program phase provided FR 5 token consequences. During the final no program phase all students' quality of performance diminished noticeably under the classroom token contingencies as compared to the tutoring program contingencies (*Reinforcement IV* and *V* phases).

In addition to the decrease in the quality of performance there was another feature of the data that was important. Concurrent with a change in the phase of the program, in particular from the tutoring setting to the classroom setting or vice versa, very sharp changes in the *level* of student performances were obtained. These sharp changes in level of performance were interpreted to be a form of discriminative behavior, and further suggested that the effects obtained with the tutoring program would not be obtained from (generalize to) other instructional settings or methods. Whether the failure of generalization of effects produced by the tutoring program was the result of extended exposure, the various manipulations of the present project were not determined. There appeared to be no reason why a program fading out (thinning) the frequencies of token and praise consequences and gradually fading in the classroom setting would not have maintained performance that was much higher quality than baseline performances. The use of the present procedures in school settings would dictate the desirability of such a fading procedure.

Attempts to evaluate contingencies effecting the tutors' quality of performance were equivocal. Part of indeterminate results may have been the result of relatively weak manipulations that were employed during the no program phases. However, it was not feasible to evaluate the various contingencies with the sixth grade students in their classroom. Therefore, all contingency changes after the first no program phase were made with the tutors out of their classroom which meant that the free time contingency was confounded with the other contingencies that were scheduled. A second feature of the manipulations that may have been weak was the choice of comparison conditions. That is, the tutoring program maintained high quality performance, and these performance levels served as the comparison or baseline levels from which the effects of the various no program contingencies were evaluated. The use of the various no program phase contingencies as comparison conditions or the extension of those phases another five or ten sessions may have resulted in different tutor data since the no program phases were clearly less preferred as indicated by the tutors' complaints about the onset of no program phases.

The scheduling of free time from the classroom contingent upon completion of the SRA assignments appeared to result in improved quality of performance (of *No Program II* - FREE TIME with *No Program I*). However, the effects of practice with the SRA materials were not clearly implicated by the present data. That is, the scheduling of the free time contingency and SRA practice (*No Program III* phase) in the third grade classroom resulted in an improvement in Curtis' performance. But, no improvement in Kevin's performance could be detected because of a ceiling effect during the previous no program phase. The possibility of a practice effect would seem reasonable since both third and sixth grade students were using the SRA materials, and those materials are formated similarly at the various skill levels and require similar types of responses at the various skill levels.

A rather baffling effect of the program was that the quality of the tutors' performance was susceptible to change during manipulations of contingencies that were placed on completing the exercises. Whether this effect (a) was specific to the SRA exercises, (b) was representative of a broader interaction between improvements in the quality of performance and when the programmed materials increase in difficulty or skill level, or (c) was related to some other unspecified variable was not determined.

While the absenteeism data appear to be related to the students' participation in the tutoring program, those data cannot be said to have *necessarily* resulted from the program. Reduced absenteeism may have resulted from factors other than the program. The fact that increases in the rate of absenteeism did not consistently occur during scheduled no program phases (see Table 4) would support a conservative interpretation of the absenteeism data. However, extending the duration of the no program phases may have resulted in consistent increases in absenteeism.

One feature of the project that was crucial to the success of the program if not contributing to the improved quality of the tutors' performance was the monitoring of the tutoring sessions. Tutoring behaviors were monitored daily, and the tutors were provided with regular feedback concerning the adequacy of their teaching responses. A less frequent schedule or some daily sampling procedure might be adequate to obtain the quality of teaching responses necessary to maintain quality student performance; however, these possibilities do need to be explored if the tutoring program is to be adapted to more general use. The present monitoring system required that the two raters each spend about fifteen hours per week evaluating the tape recorded tutoring sessions. By not having daily rater reliability checks that amount of time could be reduced, but the program should be expanded to handle a larger number of students which would mean more monitoring of tutors.

REFERENCES

Bailey, J. S., Piotrowski, W. and Johnson, M. The elementary school student as tutor in reading and arithmetic. Paper read at the Third Annual Kansas Conference on Behavior Analysis in Education, Lawrence, Kansas, 1972.

Conlon, M. F., Hall, C. and Hanley, E. M. The effects of a peer correction procedure on the arithmetic accuracy for two elementary school children. Paper read at the Third Annual Kansas Conference on Behavior Analysis in Education, Lawrence, Kansas, 1972.

Davis, M., Morris, J. and Price, G. Token reinforcement programs in the public school. Paper presented at the 1970 meetings of the Rocky Mountain Psychological Association, Salt Lake City, Utah.

Graubard, P. S., Rosenberg, H. and Miller, M. M. Student applications of behavior modification to teachers and environments or ecological approaches to social deviance. In E. A. Ramp and B. L. Hopkins (Eds.), *A new direction for education: behavior analysis 1971*. Lawrence, Kansas: The University of Kansas Support and Development Center for Follow Through, 1971, pp. 80-101.

Hamblin, R. L., Hathaway, C. and Wodarski, J. Group contingencies, peer tutoring, and accelerating academic achievement. In E. A. Ramp and B. L. Hopkins (Eds.), *A new direction for education: behavior analysis 1971*. Lawrence, Kansas: The University of Kansas Support and Development Center for Follow Through, 1971, pp. 41-53.

Harris, V. W., Sherman, J. A. and Henderson, D. G. The effect of a tutoring procedure on the spelling performance of elementary classroom students. Paper read at the Third Annual Kansas Conference on Behavior Analysis in Education, Lawrence, Kansas, 1972.

Madsen, C. H. and Madsen, C. K. *Teaching/discipline: behavioral principles toward a positive approach*. Boston: Allyn and Bacon, 1970.

Parker, D. H. and Scannell, G. *SRA reading laboratory*. Chicago: Science Research Associates, 1961.

Surratt, P. R., Ulrich, R. E. and Hawkins, R. P. An elementary student as a behavioral engineer. *Journal of Applied Behavior Analysis*, 1969, 2, 85-92.

Willis, J. W. and Crowder, J. A behavioral approach to remedial reading using students as behavioral engineers. Paper read at the Third Annual Kansas Conference on Behavior Analysis in Education, Lawrence, Kansas, 1972.

Part 5:

Programs & Techniques in Behavior Analysis:

teacher training

SETTING CRITERION RATES OF TEACHER PRAISE: THE EFFECTS OF VIDEO TAPE FEEDBACK IN A BEHAVIOR ANALYSIS FOLLOW THROUGH CLASSROOM[1,2]

richard a. saudargas[3]

Two components are usually present in training teachers in behavior modification. The first consists of the introductory material describing the principles of operant theory, how the theory is applied to the classroom, and either implicit or explicit instructions as to how teachers should behave in the classroom. The second component involves feedback to the teacher about his or her performance in the classroom. Research has indicated that specific instructions do not lead to reliable changes in teachers' classroom behaviors (Buel, 1970; Gelfand, Elton and Harmon, in press; Madsen, Becker and Thomas, 1968). Feedback procedures, however, have been found to be relatively effective in changing teachers' behaviors for at least as long as the feedback procedures are in effect (Bricker, Morgan and Grabowski, 1970; Cooper, Thomson and Baer, 1970; Panyon, Boozer and Morris, 1970; Cossairt, Hall and Hopkins, 1972; Thomas, 1972).

One problem with giving teachers classroom feedback concerns how the feedback is transmitted. In one study, feedback was given by the classroom observers (Cooper et al., 1971). In a second study, Thomas (1972) used videotape recordings to provide teachers with feedback. Neither of these studies reported data which indicated whether teachers behaved appropriately in the absence of the observers or the videotape recorders.

---

[1] The present paper is based on a dissertation submitted to the Department of Psychology, The Florida State University, in partial fulfillment of the requirements for the Ph.D. degree. The author would like to thank Drs. Charles Madsen, Jr., Jon Bailey, Jack Hokanson, George Weaver and Edwin Smith for their suggestions during the course of the research. Thanks are also extended to Dr. Don Bushell, Jr., who made the research possible.

[2] The project presented was supported by a grant to Dr. Don Bushell, Jr., from the U.S. Office of Health, Education and Welfare (G-0-8-522422-433[286]).

[3] Now at the University of Kansas, Department of Human Development and Family Life.

The present study analyzed the effect of videotape feedback on establishing criterion rates of teacher praise. In order to determine the discriminative effects of the presence of the videotape recorder, additional data were obtained by observers when the videotape recorders were not in operation.

METHOD

Setting

The study was conducted in one first grade and one third grade class in Philadelphia, Pennsylvania. Children who attended the training center were area residents.

The classroom day was divided into a series of alternating "earn periods" and "spend periods." During each 30-minute earn period children earned tokens for working in one of four assigned academic areas: reading, handwriting and spelling, mathematics, and language arts. The children rotated through each academic area at least once a day and each child was allowed to work at his own pace. Each earn period was followed by a 20-minute spend period during which children exchanged tokens for games and other reinforcing activities.

Teachers

One first grade teacher aide and one third grade teacher aide who taught handwriting and spelling served as subjects. Both aides were mothers of children in the Behavior Analysis Follow Through program.

Daily Data Collection Procedures

VIDEOTAPE RECORDING. During the first morning earn period, each teacher was videotaped for ten minutes. Observers rated the videotapes in the afternoon.

OBSERVER RECORDING. Teacher A was observed for seven minutes during each of the three morning earn periods when the videotape recorder was not in operation. Teacher B was observed during the first two morning earn periods. Observer recorded between the 10th and 20th minute of each earn period. Approximately one hour separated each observation.

MONITORING TEACHER PRAISE RATES (Academic Approvals). Academic Approvals were defined as praise statements made by the teacher to a specific child contingent upon correct student academic behavior. Behaviors such as raised hands and sitting quietly were not considered academic because the child could emit them in the absence of any books.

Three parents of community children were trained to record the occurrence of Academic Approvals. Observations of teacher-child interactions were continuously recorded in discrete one minute time intervals. Two types of reliability were calculated. Occurrence reliability was calculated between two observers and determined the extent to which the observers recorded the Academic Approvals at the same time. Occurrence reliability was obtained by dividing the number of agreements by the number of agreements plus disagreements between the observers (Thomas, Becker and Armstrong, 1968). At least two reliability checks were made during each of the conditions for observer recorded data. Occurrence reliability for Teacher A varied between .76 and 1.00 with a mean of .86. The mean occurrence reliability for Teacher B was .87 and ranged between .71 and 1.00. Frequency reliability was calculated between the teachers' frequency counts and the observers' frequency counts taken from the videotape recordings. Frequency reliability was obtained by dividing the lower frequency by the higher frequency for the ten minute videotape recording. Frequency reliability was calculated daily when the teachers counted Academic Approvals. Mean frequency reliability for Teacher A was .88 (range = .81 - 1.00) and for Teacher B, .87 (range = .76 - .96).

Experimental Conditions

BASELINE. No manipulations of teacher behavior were attempted during the baseline conditions. The teachers were monitored by both videotape recorders and observers.

COUNT + GRAPH + CRITERION RATE OF ACADEMIC APPROVALS. Each teacher was asked to meet a specified criterion rate of Academic Approvals. On a graph posted near the teacher's instructional area, a line was drawn so that the teacher could see whether or not the criterion had been met. Two different criteria were set. One criterion was set at 5.0 Academic Approvals per minute. The second was established at 2.0 Academic Approvals per minute.

Each teacher was trained separately to count and graph Academic Approvals rates from her own videotape recordings on the first day of the first experimental condition. No additional assistance was given to the teachers after the first day of training. Teacher ratings of the videotape recordings were made each day in the afternoon about four hours after the videotapes were made. After calculating her rate, she plotted it on her classroom graph.

During the last session of each experimental condition, the teachers were told that they would no longer be required to observe their videotape recordings and that there were no restrictions on rates of Academic Approval.

Each criterion rate condition was run until the teacher's rate met or exceeded criterion for at least two consecutive observation days for the observer recorded data of Earn Period I. Baseline conditions were changed when the baseline rate was not within the criterion range for at least two consecutive observation days for the observer recorded data of Earn Period I.

RESEARCH DESIGN. A reversal design was used with the experimental conditions sequences in opposite order for the two teachers.

The research design for Teacher A was as follows:

- *Baseline 1*
- *Count + Graph + 5.0 per minute criterion*
- *Baseline 2*
- *Count + Graph + 2.0 per minute criterion*
- *Count + Graph + 5.0 per minute criterion*
- *Count + Graph + 2.0 per minute criterion*

The research design for Teacher B was as follows:

- *Baseline 1*
- *Count + Graph + 2.0 per minute criterion*
- *Baseline 2*
- *Count + Graph + 5.0 per minute criterion*
- *Count + Graph + 2.0 per minute criterion*
- *Count + Graph + 5.0 per minute criterion*

RESULTS

Rates per minute were graphed for each teacher and median lines were drawn through the data points for each experimental condition.

Teacher A (Figure 1)

The rate data for Baseline 1 were stable during the observed time periods. The medians ranged from 2.6 to 3.2 Academic Approvals per minute. When the criterion was set at 5.0 Academic Approvals per minute, daily trends in the direction of the criterion were observed during all time periods. The median rate for the Videotape Monitored data was above the criterion.

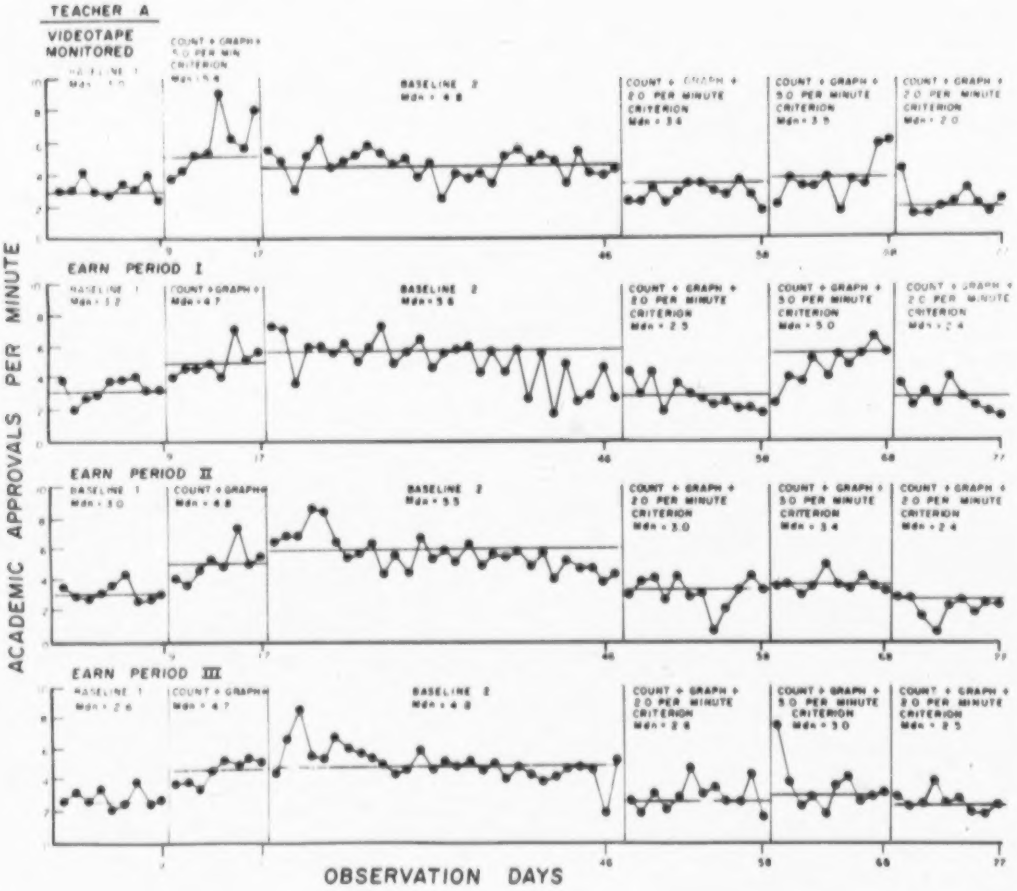


Fig. 1. Teacher A rates of Academic Approvals for each observation period.

During Baseline 2 the rates were higher than Baseline 1 in all observation periods. Rates during Earn Period I and Earn Period II appeared to decline more than the Videotape Monitored and Earn Period II data. When the criterion was shifted to 2.0 per minute, the rate of Academic Approvals decreased toward the criterion value in all observation time periods.

When the criterion was returned to 5.0 per minute, the observer recorded data of Earn Period I showed an almost daily increase toward the criterion rate. The Videotape Monitored data showed an upward trend only for the last two days of the condition. There did not appear to be any differences in rate from the previous 2.0 criterion condition during Earn Period II and Earn Period III.

During the second 2.0 condition, the rate of Academic Approvals for the Videotape Monitored time period decreased to criterion and stabilized around 2.0 per minute. The observer monitored data of Earn Period I decreased to a rate slightly below the criterion rate. It was difficult to discern any rate changes for the observer data recorded during Earn Period II and Earn Period III.

Teacher B (Figure 2)

The median rates for Baseline 1 varied within 1.0 Academic Approvals per minute between observation times, but there were two noticeable trends in the data. One trend appeared on day 7 during the Videotape Monitored condition when rate increased sharply. The data then de-

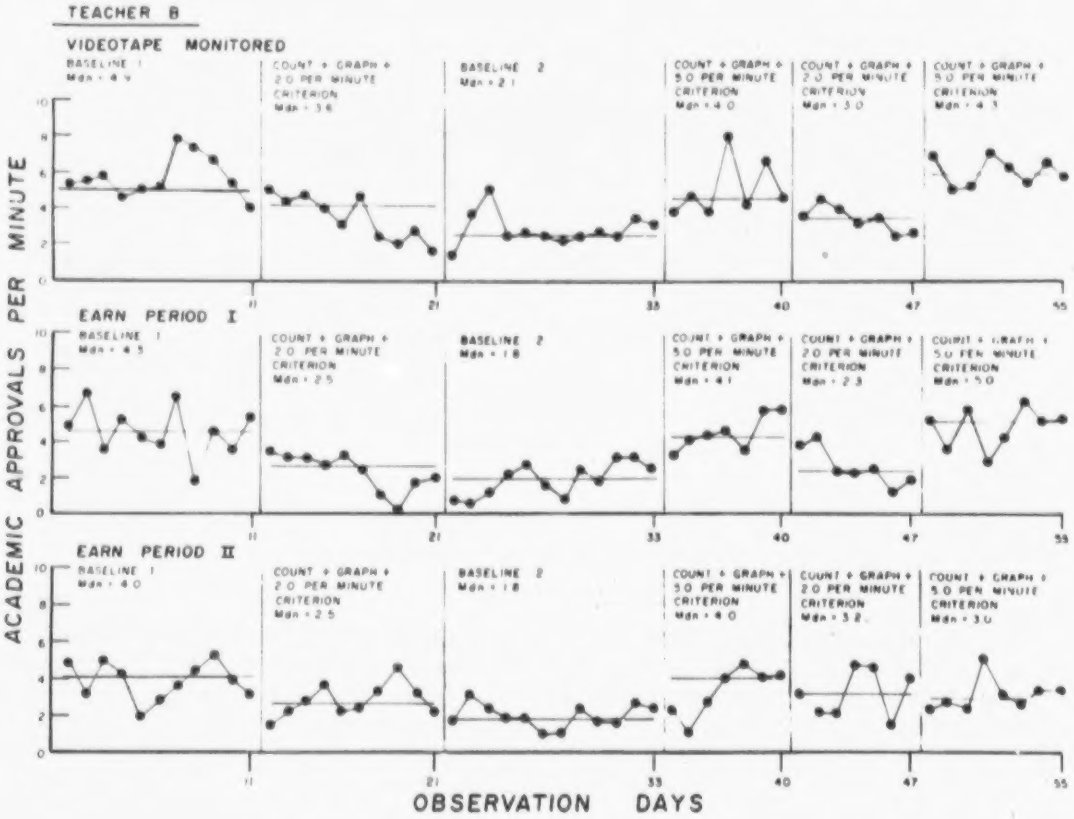


Fig. 2. Teacher B Academic Approvals for each observation period.

creased steadily until the end of the baseline condition. The data for Earn Period I did not appear to follow the trend of the Videotape Monitored data. However, during Earn Period II an apparently decreasing baseline reversed after day 5 and increased steadily until day 9.

During the first criterion condition (2.0 Academic Approvals per minute), the rates for the Videotape Monitored data and Observer Monitored data for Earn Period I moved toward the criterion rate. There was an overall decrease in rate of Academic Approvals during Earn Period II.

During Baseline 2, rates remained relatively low and stable during all observation time periods.

When the criterion was changed to 5.0 Academic Approvals per minute, the rates of Earn Period I and Earn Period II increased from the Baseline 2 to the criterion rate. There was a median increase from Baseline 2 for the Videotape Monitored observations but no general tendency to meet the criterion rate.

As in the previous criterion 2.0 condition, both the Videotape Monitored observations and the data recorded during Earn Period I followed the same general trend in the direction of the criterion rate. The rates of Earn Period II were highly variable.

Upon returning to a criterion rate of 5.0 Academic Approvals per minute, an increase in observed rates appeared evident for the three observation time periods. Only the data recorded for Earn Period I met the minimum stability criterion.

DISCUSSION

The data indicated that having teachers count, graph, and attempt to meet two different criterion rates of Academic Approvals effectively increased and decreased teacher rates. The data further indicated that the criterion rate changes were produced most reliably during the time period the teachers were being videotaped and during the observation time period immediately following the videotape recording. The results recorded for Earn Periods later in the morning were not consistent within or between teachers. In general, the effect of setting criterion rates was less evident during Earn Periods II and III (Teacher A), and Earn Period II (Teacher B), especially during the last two experimental conditions.

The day to day tendency for the rates to move in the direction of the criterion rate was relatively consistent between teachers. One teacher commented, however, that pacing her praise during the criterion conditions was difficult to do. The comment by the teacher suggests a timing which had to be learned each time the criterion rate was changed.

A second purpose of the present experiment was to determine if teachers would discriminate the feedback transmitter. If there was a discriminative effect, the teachers would have behaved differentially in the presence of the videotape recorder and the observers. The results of this study suggest little if any discrimination between the videotape monitoring and the observer recording during Earn Period I. The changes in rate brought about by changing the criterion requirements were observed by both the videotape recorders and the observers.

The teachers were observed to have emitted Academic Approvals at the rate of the criterion condition prior to Baseline 2. The median rate of Teacher A during Baseline I was 3.0 Academic Approvals per minute. When she had experienced meeting a criterion rate of 5.0, the rate of Baseline 2 was 4.8 per minute. For Teacher B the Baseline I rate of Academic Approvals was 4.8 per minute. The median rate for Baseline 2 after experience meeting a criterion rate of 2.0 per minute was 2.1 Academic Approvals per minute. These results suggest that the teachers' behavior came under the control of reinforcing stimuli other than the videotape feedback. Future research should be done to replicate and clarify these findings.

In the present experiment the videotape recordings were observed each day the criterion conditions were in effect. One question raised by others (e.g., Thomas, 1972) is concerned with the feedback schedule which would maintain a desired rate of teacher behavior. An extension of the present experiment would be to vary the schedule with which the teachers receive feedback.

One aspect of generality is concerned with applying a variable to different behaviors (Sidman, 1960). The present techniques could easily be applied to other teacher behaviors such as questioning, disapproving, and long chains of teaching behavior. If it could be shown that these techniques could also be used to change behaviors other than praise, a greater degree of generality would be demonstrated.

It might not be necessary to use videotape feedback if other mechanisms could be found such as trained observers or supervisors. However, other research which has used observers and supervisors as feedback transmitters has been equivocal (Cossairt et al., 1972; Thomson, Holmberg and Baer, 1972). None of the previous research on feedback to teachers has set criterion rates. Setting a criterion may be critical for consistent changes in teacher behaviors.

REFERENCES

Bricker, W. A., Morgan, D. and Grabowski, J. Token reinforcement of attendants who work with low functioning children. *Abstracts of Peabody Studies in Mental Retardation*, 1965-1968, 4, 1968.

Buel, C. L. Effects of instructions, monetary bonus contingencies, and training on behavior modifying skills of a nursing staff. Unpublished master's thesis, University of Utah, 1970.

Cooper, Margaret L., Thomson, Carolyn L. and Baer, D. M. The experimental modification of teacher attending behavior. *Journal of Applied Behavior Analysis*, 1970, 3, 153-157.

Cossairt, A., Hall, R. V. and Hopkins, B. L. The effects of experimenter's instructions, feedback, and praise on teacher praise and student attending behavior. Paper presented at Kansas Conference on Teacher Training, Lawrence, Kansas, 1972.

Gelfand, Donna, M., Elton R. H. and Harmon R. E. A videotape feedback training method to teach behavior modification skills to nonprofessionals. *Journal of Applied Behavior Analysis*, in press.

Madsen, C. H., Jr., Becker, W. C. and Thomas, D. R. Rules, praise, and ignoring: Elements of elementary classroom control. *Journal of Applied Behavior Analysis*, 1968, 2, 139-150.

Madsen, C. H., Jr. and Madsen, C. K. *Observational and learning activities manual for teaching discipline: Behavioral principles towards a positive approach*. Boston: Allyn and Bacon, 1973.

Panyon, Marion, Boozer, H. and Morris, N. Feedback to attendants as a reinforcer to applying operant techniques. *Journal of Applied Behavior Analysis*, 1970, 3, 1-4.

Sidman, Murray. *Tactics of scientific research*. New York: Basic Books, 1960.

Thomas, D. R. Self-monitoring as a technique for modifying teaching behaviors. Unpublished doctoral dissertation, University of Illinois, 1972.

Thomas, D. R., Becker, W. C. and Armstrong, Marianne. Production and elimination of disruptive classroom behavior by systematically varying teacher's behavior. *Journal of Applied Behavior Analysis*, 1968, 1, 35-45.

Thomson, C. L., Holmberg, M. C. and Baer, D. M. The experimental analysis of training procedures for preschool teachers. Paper presented at the Kansas Conference on Teacher Training, Lawrence, Kansas, 1972.

SCHEDULES OF FOLLOW-UP AND THEIR EFFECT UPON THE MAINTENANCE OF A PRESCRIPTIVE TEACHING PROGRAM

beverley holden
beth sulzer-azaroff

INTRODUCTION

The purpose of this study was to determine how a special class teacher could conduct an effective and efficient follow-up program with students placed back in regular classrooms. The relationships between two different schedules with which the special teacher contacted the regular classroom teacher were compared to assess their effect upon the students' subsequent performance.

METHOD

Subjects

The subjects in this study were five children and their regular class teachers. The children were enrolled in the Pupil Assessment Class at Southern Illinois University for a period of at least five weeks. All were first graders. The children had been referred to this class because of various learning problems in the areas of visual discrimination and spatial organization. The special classroom teacher (E) was in charge of the academic program. The regular classroom teachers were those who had taught the child both prior to his entrance in the special program and upon his departure from it.

Experimental Space, Apparatus and Materials

The children were enrolled in the Assessment Class program before returning to their regular classroom. The class was equipped with furnishings similar to those found in a typical public school. At least two hours of the day in the room were spent in group instruction of reading, arithmetic and language. The other two hours were spent in individual activities designed to determine the children's difficulties and how to remediate them.

Specific activities used by the subjects included a series of visual discrimination exercises (Task I): Marianne Frostig Sheets, Scott Foresman Open Highways sheets, and exercises from the Continental Press.

The Fitzhugh Spatial Organization Series -- *101*, *102*, and *104* (1968) constituted Task II. Both tasks consisted of pages containing either four or six individual frames in which the requirement was to note similarities and differences by circling appropriate items. Responses were defined as right or wrong.

Reliability

Reliability of scoring presented no problem since there was a permanent record in the form of the student's worksheets. Worksheets were compared with the teacher's data sheets. In every case there was 100% agreement.

Procedure

When the children returned to the regular classroom they took a prescriptive teaching program with them. This included lesson ideas in the weak areas and worksheets as described above for remediation. The regular teacher was asked to write the student's name and the date at the top of each paper the child completed. She was also given a sheet to fill out each day. On it she marked the exercises the child did and the number of correct responses. A folder was provided for storing the student's work. She was asked to save the sheets so that the E could collect them at the end of the project.

The telephone was used as the means for follow-up. Two phone call schedules were used to control for order effects. Order I consisted first of a baseline. During this time the teacher was not called until the last day of the two-week period. Next, there was an FI 5 schedule, during which the teacher was called every fifth school day for two weeks. The next phase was an FI 1, during which the teacher was called every school day for two weeks. The next step was a return to FI 5 for two weeks, then a return to FI 1 for the same amount of time. Order II consisted of the reverse of the above. First, was the two week baseline, next the FI 1 schedule, then the FI 5 schedule for two weeks, then a return to the FI 1, and finally to the FI 5 schedule again. (See Table 1.)

When the special class teacher called the regular class teacher she only asked about one of the two assigned tasks. This was done to determine if inquiring about one task only would produce a differential effect upon performance on each of the two tasks (i.e., would inquiring about Task I produce a relatively higher rate of performance on that task and vice versa). Usually the special teacher would ask how things were in general in the classroom and then would say that she wanted to find out how the child was doing with one task. If the regular teacher made mention of the other task the special teacher would usually change the subject and return to the one selected for inquiry. She would also give the teacher verbal reinforcement when appropriate, by making some comment such as, "I'm certain-

TABLE 1

FOLLOW-UP PROCEDURE

SUBJECT	TASKS ASSIGNED	TASK INQUIRED ABOUT	ORDER OF PHONE CALL SCHEDULES CHANGED EVERY TWO WEEKS
1	FITZHUGH & VISUAL DISCRIMINATION	VISUAL DISCRIMINATION	BASELINE FI 5, FI 1, FI 5, FI 1
2	FITZHUGH & VISUAL DISCRIMINATION	FITZHUGH	BASELINE FI 5, FI 1, FI 5, FI 1
3	FITZHUGH & VISUAL DISCRIMINATION	VISUAL DISCRIMINATION	BASELINE FI 1, FI 5, FI 1, FI 5
4	FITZHUGH & VISUAL DISCRIMINATION	FITZHUGH	BASELINE FI 1, FI 5, FI 1, FI 5
5	FITZHUGH & VISUAL DISCRIMINATION	VISUAL DISCRIMINATION	BASELINE FI 5, FI 1, FI 5, FI 1

ly glad that you are willing to keep in contact with me and are helping with ______________ (child's) work."

After the calling phases of the program were completed, the E collected the student's work folder by making a visit to the school and continued to keep in contact at least once a month for the remainder of the school year.

RESULTS

For summary purposes the combined results of *Subjects 1* through *4* are shown in Figures 1 and 2. *Subject 5's* results were not included since very little data were obtained on that *S*.

The total number correct and completed in both the E-inquired and uninquired tasks are shown in Figures 1 and 2. These show comparatively higher rates of performance when phone calls were made each day rather than once a week. Thus, the more phone calls the more work the child accomplished.

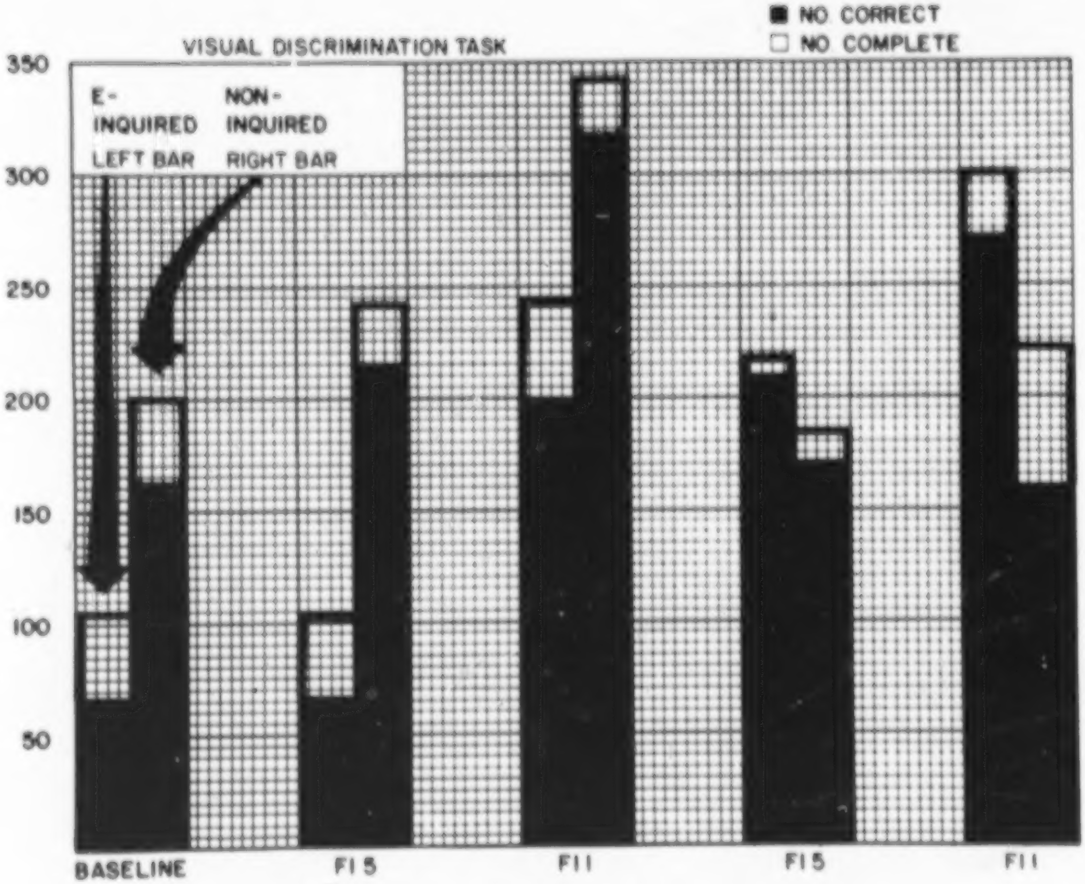


Fig. 1. Cumulative number of correct and completed responses for four subjects on the visual discrimination task.

In the visual discrimination task, initially there was a higher performance rate on the uninquired task than the inquired one but as time progressed the rate for the inquired work increased to a much higher proportion, particularly in the last two weeks of the study. With the Fitzhugh task the rate on the E-inquired task was higher in four of the phases than when it was not inquired about.

Both the cumulative number completed and the number correct were tabulated for each *S*, under each feedback condition. Results for each *S*'s performance on each task are shown in Figures 3 through 11. Figures 3 and 4 show the work of *Subject 1* for whom the E-inquired task was the

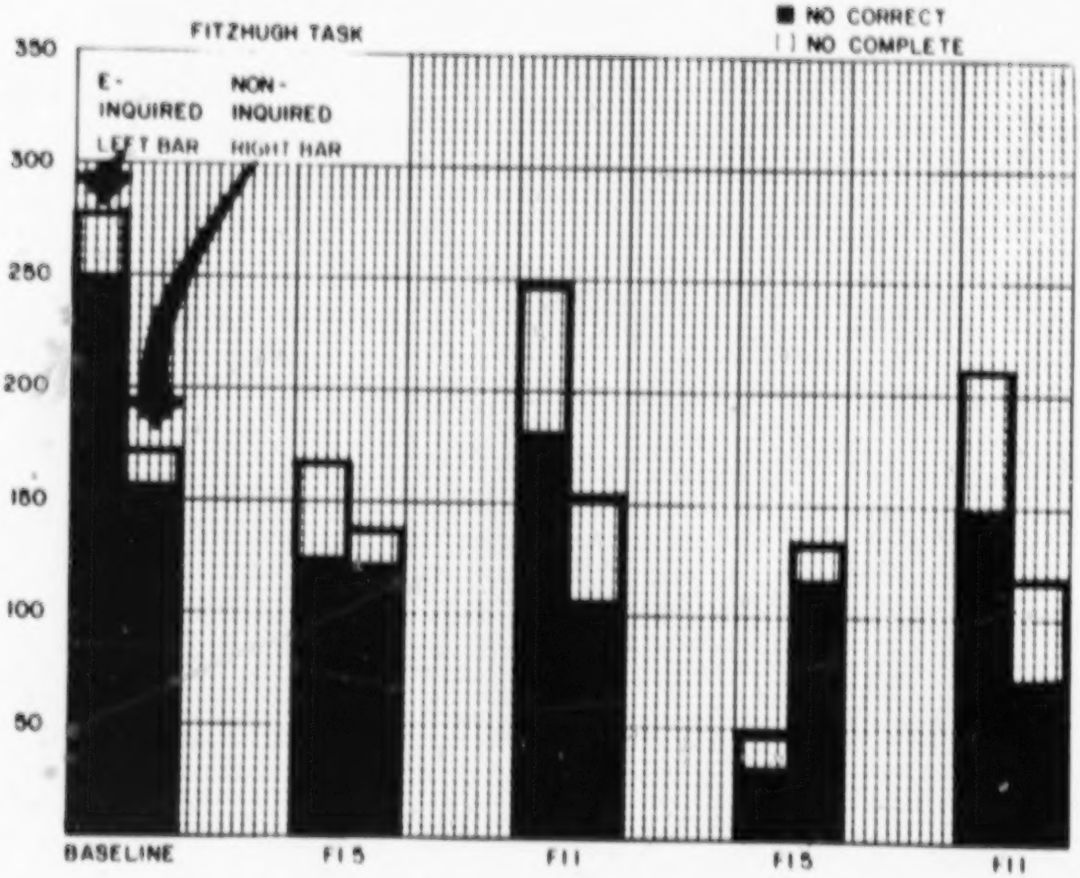


Fig. 2. Cumulative number of correct and completed responses for four subjects on the Fitzhugh task.

visual discrimination one. During baseline the teacher had the child do the work every day except the last. In the FI 5 phase the cumulative performance dropped slightly. In the third phase (FI 1) the cumulative rate of responses rose subtantially from the prior phase. When weekly calls resumed, the rate dropped to 120 from 174 correct during the previous phase. When phone calls were again increased to one each day, the rate rose for the number completed in the E-inquired task but dropped slightly for number correct.

E did not inquire about the Fitzhugh task for the subject shown in Figures 3 and 4. However, the classroom teacher still did the work with the child. The rate dropped slightly during the FI 5 phase and then rose in the FI 1 phase. The rate dropped when phone calls were again

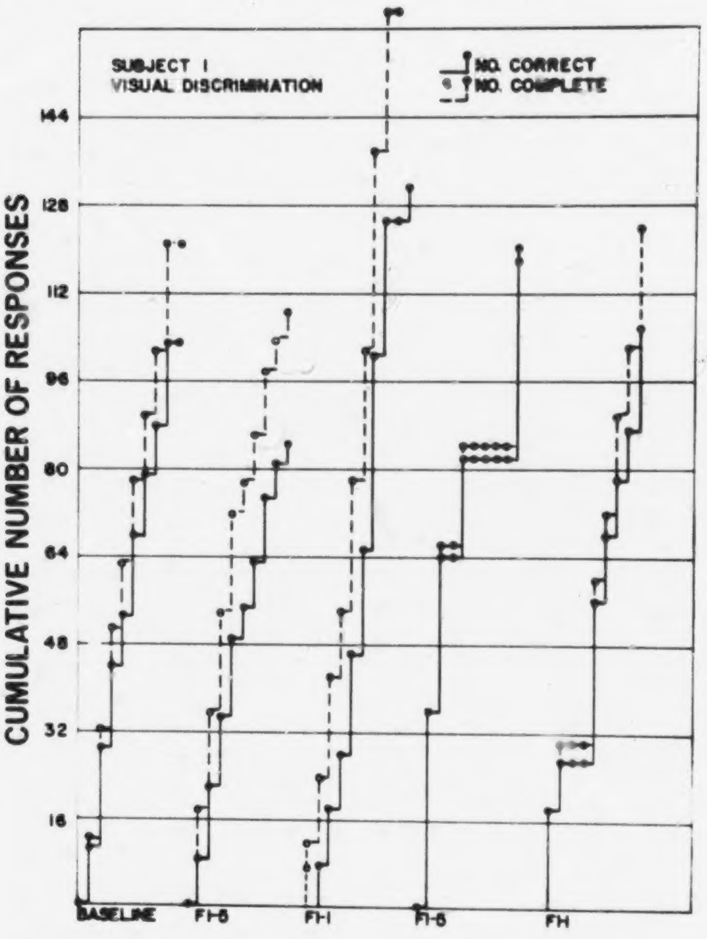


Fig. 3. Subject I, Visual Discrimination Task.

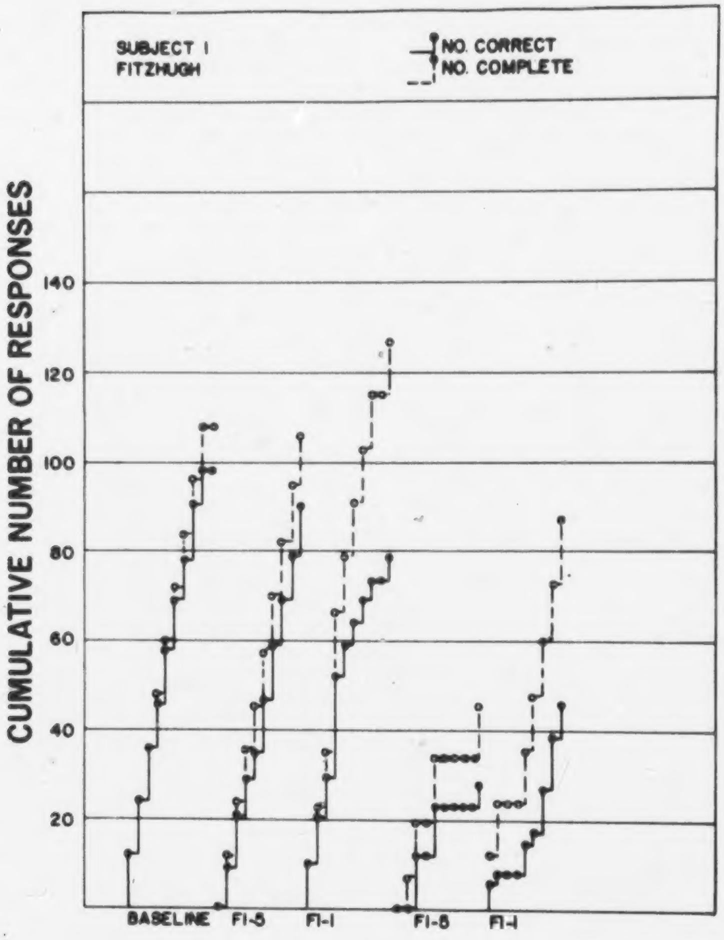


Fig 4. Subject I, Fitzhugh Task.

only made once a week. Another increase was seen when phone calls were again made every day. The number completed rose from 46 to 87 while the number correct rose from 28 to 46.

Comparing the tasks, there was a higher rate of work for the E-inquired task than for the other (Fitzhugh). The difference is most apparent in the second FI 5 schedule.

Figures 5 and 6 show the performance of *Subject 2* for whom the E-inquired task was the Fitzhugh. During baseline, 108 Fitzhugh items were completed, 88 of them correct. In the FI 5 phase, the cumulative number correct dropped but the number completed stayed the same. When calls were made each day the rate rose but only four points for the number completed and eight points for the number correct. With calls again occurring only once a week, the number completed dropped from 112 to 47 and

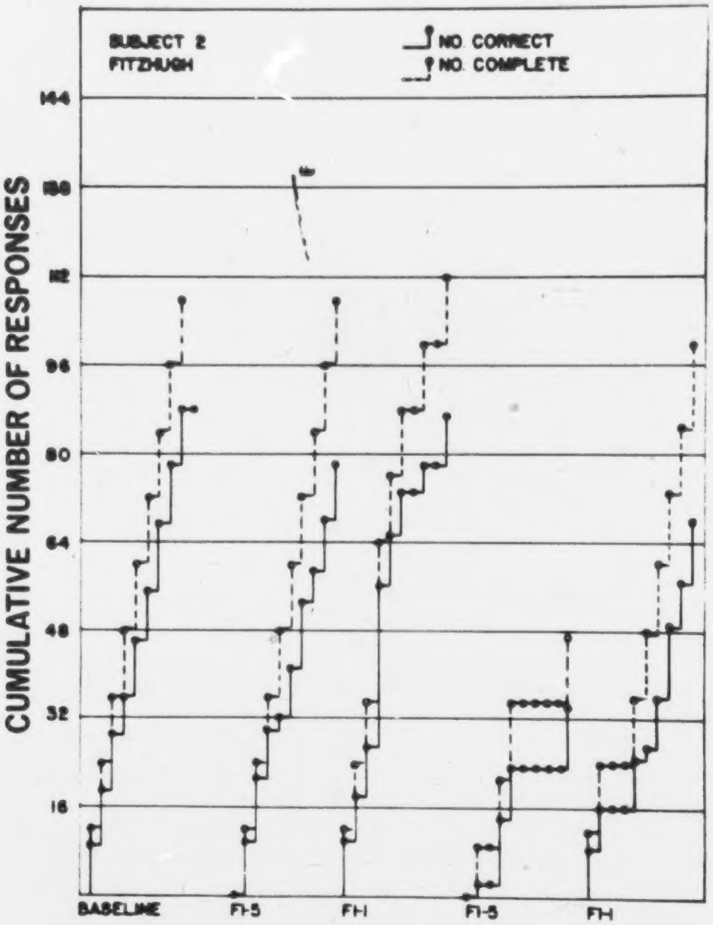


Fig. 5. Subject 2, Fitzhugh Task.

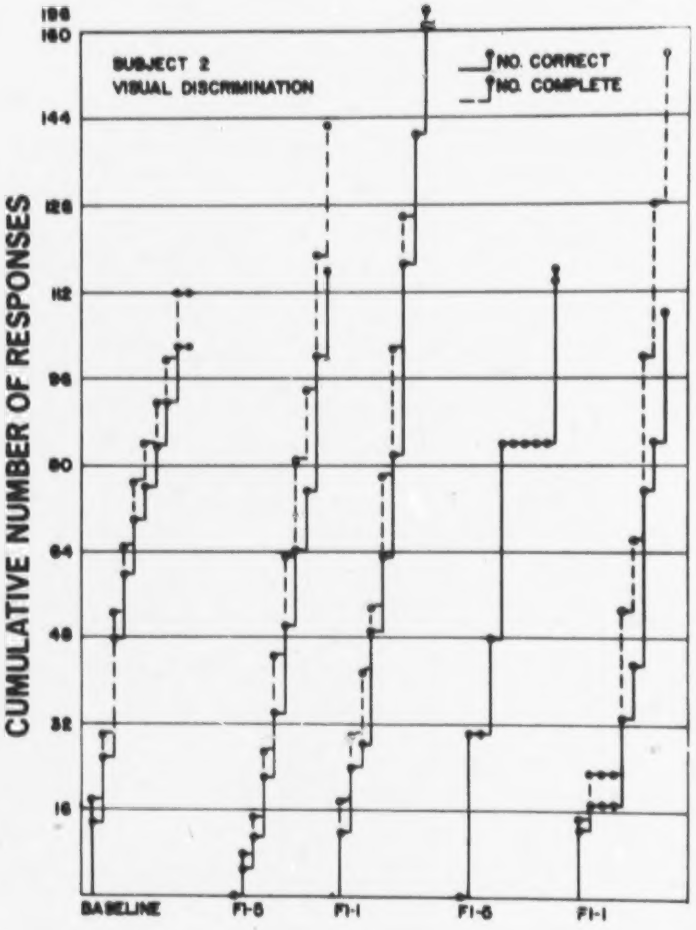


Fig. 6. Subject 2, Visual Discrimination Task.

the number correct dropped from 86 to 34. With the return to FI 1 the rates for number correct and completed rose.

In the uninquired visual discrimination task the baseline rate for number correct and completed was 112. During the next phase (FI 5) the number correct and completed rose slightly. With phone calls made each day the number correct and complete continued to rise. When phone calls were made only weekly during the second FI 5 phase, there was a drastic drop in both the number correct and completed. When the teacher was again called each day, the cumulative rate number completed rose but the number correct did not.

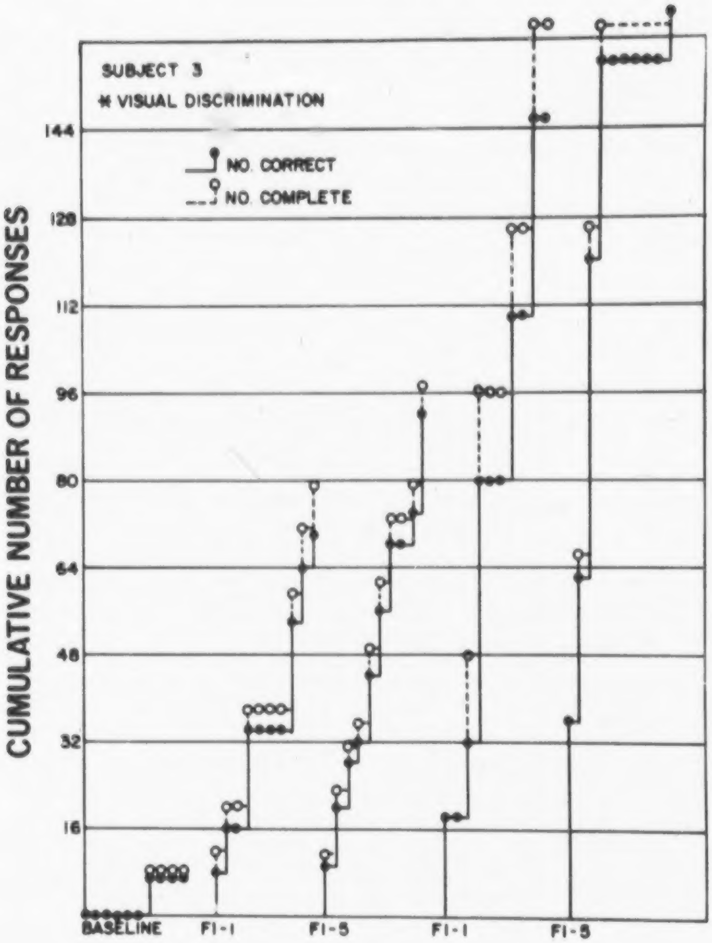


Fig. 7. Subject 3, Visual Discrimination Task.

The results of the work of *Subject 3* are recorded on Figures 7 and 8. The baseline for the visual discrimination task (E-inquired) was very low. The S̲ did very few sheets of work during this time. During the FI 1 phase there was a substantial increase in the amount of work the child did in this task. The number completed rose to 79; the number correct 70. During the next phase (FI 5) the rate of work both completed and correct rose. It continued to rise during the reinstatement of the FI 1 phase and reached its highest rate when telephone calls were made only once a week. Thus, there was a steady increase in work completed and correct in the visual discrimination task during the progress of the project.

In the Fitzhugh task, the uninquired one, S̲ did more of the work during the baseline than during the following three phases. When the FI 1 was instated the work completed and correct dropped from 62 to 27. During

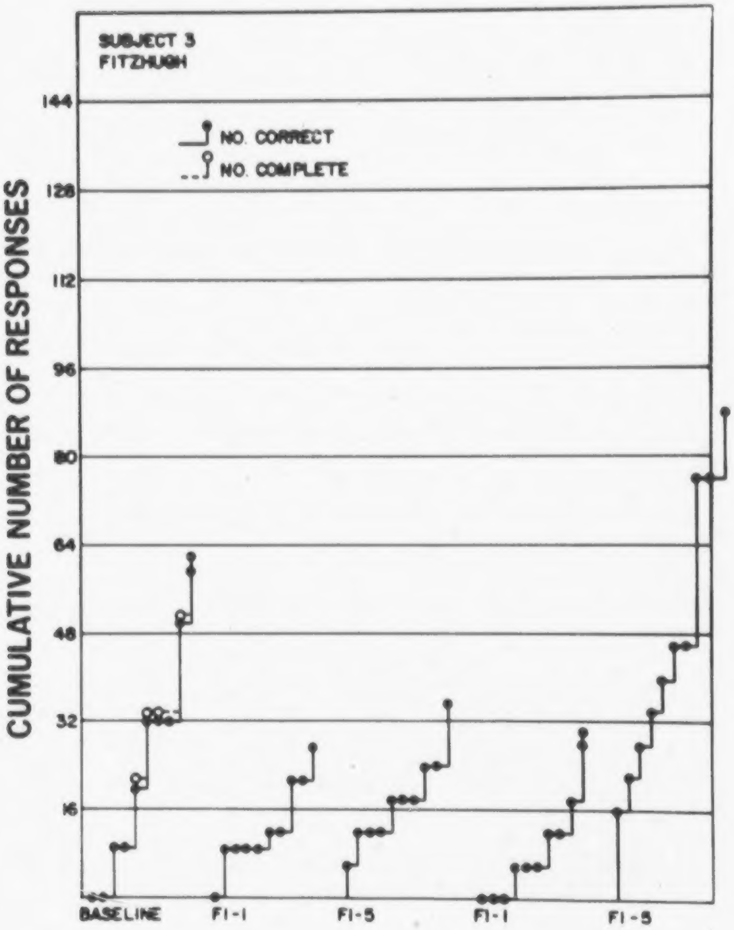


Fig. 8. Subject 3, Fitzhugh Task.

the FI 5 phase the amount of work rose but only 8 points. When daily phone calls were again reinstituted the rate of work completed and correct dropped 7 and 5 points respectively. Again, with weekly calls the rate rose to its maximum level.

Comparing performance on two tasks, one sees that the rate for the E-inquired task rose consistently throughout the project. The rate for the uninquired one, however, dropped after baseline and only rose again in the last phase.

Figures 9 and 10 show the results of *Subject 4's* performance. For the Fitzhugh task, the E-inquired one, during baseline the number completed and correct was 168 and 159 respectively, the highest rates in this task during the entire project. The rate decreased when contact was made with the teacher each day but it was higher than it was during the subsequent phase (FI 5). The rate then increased noticeably when calls were made each day again. During the last FI 5 the rate decreased to 0 for all ten days. The rate of work was higher in both FI 1 phases than in FI 5.

Baseline performance in the visual discrimination task (uninquired) was 89 completed and 77 correct. When daily phone call inquiries were made about the other task, the number completed rose to 146 while the number correct rose to 123. When phone calls were made only once a week both the rate completed and correct dropped; almost 50 fewer items were completed. When the daily schedule was again instituted the rates rose to their highest level and when the weekly phone calls were resumed, the rate dropped almost 100 points in both number completed and number correct.

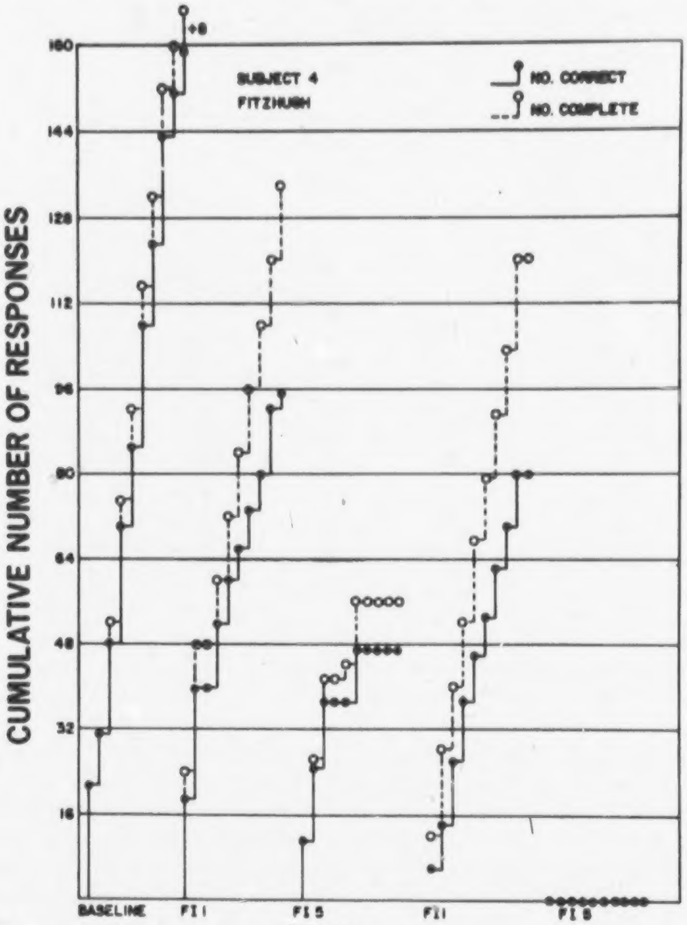


Fig. 9. Subject 4, Fitzhugh Task.

Thus, the rate in both tasks was higher in the FI 1 phases than in the FI 5 phases. The most notable difference in the Fitzhugh task was between the FI 1 phase and the FI 5 when the work dropped to 0.

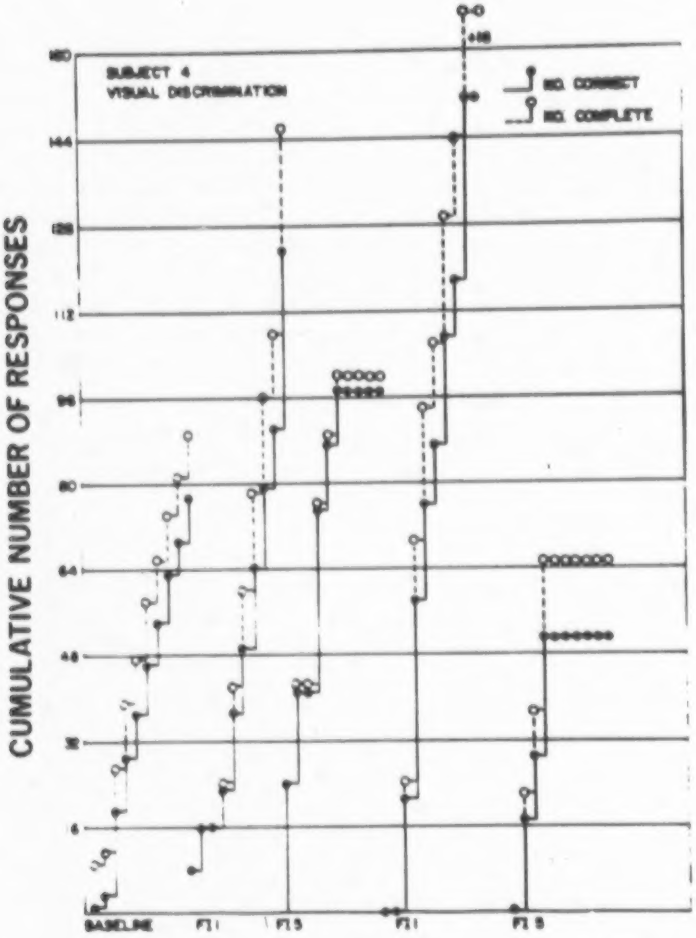


Fig. 10. Subject 4, Visual Discrimination Task.

The only recorded data for *Subject 5* are shown on the following page in Figure 11. Telephone calls were ended abruptly on the third day of the FI 1 phase. When E made the phone call, the classroom teacher asked her not to call anymore because it was too much bother for the secretary to come to get her and for her to come to the phone. E then went to collect the completed worksheets. *Subject 5's* teacher had failed to fill out the data sheets but did place dates on the visual discrimination sheets. The Fitzhugh work was not dated, although about 300 responses with the child had been completed. *Subject 5* worked on the visual discrimination task on Days 1, 2, and 5 during the baseline and no further work was done.

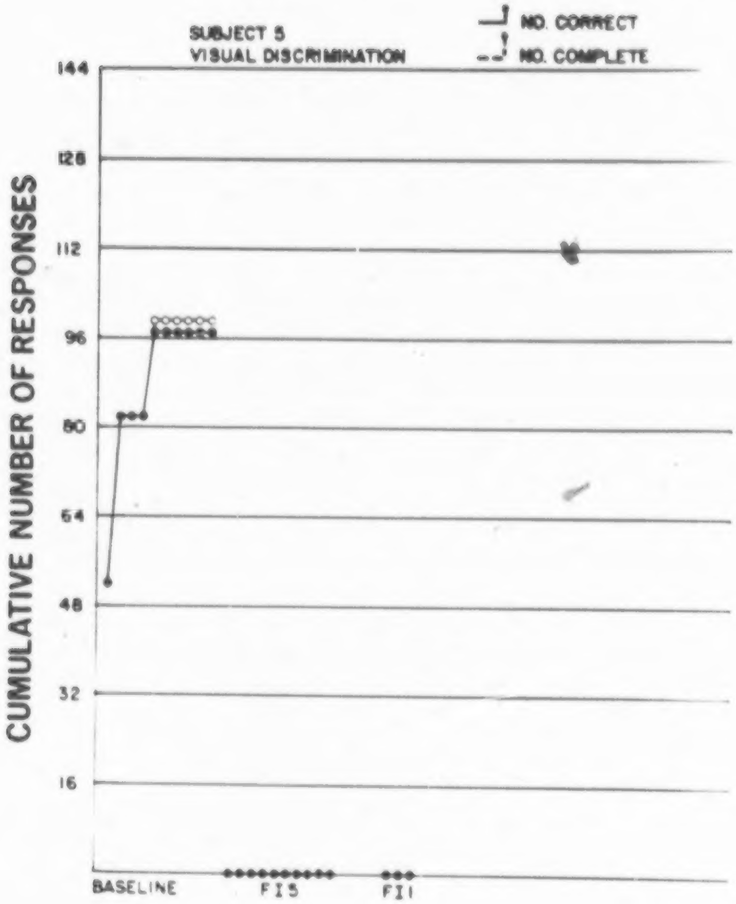


Fig. 11. Subject 5, Visual Discrimination Task.

DISCUSSION

The data suggest that telephone contact can be an effective follow-up technique for maintaining a classroom teacher's compliance with a prescriptive teaching program. Four of the teachers in this study carried out the prescriptive program at a fairly high rate. The teachers continued to see that both tasks were carried out whether the particular task was inquired about or not.

Daily phone calls were more effective for three of the subjects (*1*, *2*, and *4*). However, whether the difference was striking enough to warrant the time and bother of such a high frequency of phone calls is questionable. It appeared that some teachers (as well as the E) became tired of the daily calls. This was particularly true in the case of *Subject 5* for whom the phone calls were apparently aversive.

The brief two-week baseline phases were insufficient to determine whether compliance with the prescription would have maintained over prolonged periods of time, with no periodic contact from the E. However, the tendency towards higher student performance rates in conjunction with denser frequencies of contact suggest that such contacts are at least somewhat functionally related to compliance with the prescriptive teaching program.

A related study (Hunt and Sulzer-Azaroff, 1972) is currently being undertaken with a group of parents of pretrainable retardates. In the ongoing study, two groups of parents are the subjects. The first group of four parents was contacted regularly since the day immediately following their receipt of the prescription. The second group was not contacted for several weeks. Four of four parents in the first group continued to comply with the prescription for four weeks, while only one of the four in the second group did so during the four-week period when they were ignored. It may, therefore, be safe to assume that rates of compliance by teachers in the present study, would have gradually diminished over time and that continued contact by phone helped to maintain their rates of responding.

More research needs to be completed to examine effectiveness of other schedules of teacher contact, for example, once a month or bi-monthly. The use of a control group would help to determine a more precise evaluation of the impact of the follow-up program.

REFERENCES

Fitzhugh, K. and Fitzhugh, L. *The Fitzhugh plus program.* Galien, Michigan: Allied Education Council, 1969.

Hunt, S. and Sulzer-Azaroff, B. Motivating parent participation in home training sessions with pre-trainable retardates. Unpublished paper, 1972.

◆

A COMPARISON OF THREE DIFFERENT TYPES OF FEEDBACK ON TEACHERS' PERFORMANCE

sarah rule

Three teaching behaviors--*praise, on-task contacts* and *off-task contacts*--were modified in nine subjects. A multiple baseline design was employed to test the effects of three procedures: instructions and experimenter feedback, videotape scoring of one's own behavior, and a direct intervention procedure in which the experimenter temporarily replaced the subject whose teaching behavior fell below criterion.

Direct intervention was most effective in changing teachers' behavior. Smaller changes in the subjects' rates of praise, on-task contacts, and off-task contacts occurred during the video scoring procedure, and no predictable change occurred in the instructions plus feedback condition.

It has been consistently reported that teachers' behaviors affect children's behaviors. In particular, behaviors to which teachers attend generally increase and behaviors which they ignore decrease. A combination of positive reinforcement and extinction or punishment has been effective in changing children's behaviors (Broden, Michell, Carter and Hall, 1970; Hall, Lund and Jackson, 1968; McAllister, Stachowiak, Baer and Conderman, 1969; Madsen, Becker and Thomas, 1968; Packard, 1970; Schutte and Hopkins, 1970; Thomas, Becker and Armstrong, 1968; Wasik, Senn, Welch and Cooper, 1969).

Because teachers' behaviors affect children's behaviors, it is important to know how to teach teachers to behave effectively. Three methods have been reported in the literature concerning behavior modification. First, instructions, probably the most frequently required method of teacher training (nearly all states require teachers to take "methods" courses in which they are told how to teach) have been reported to have little effect on changing teacher behavior (Madsen, Becker and Thomas, 1968). Second, feedback to teachers about their own behavior has been demonstrated to be effective in changing their behaviors (Cooper, Thomson and Baer, 1970). Third, videotape scoring, in which teachers score tapes of themselves teaching, has also been reported to be effective in modifying the behavior of teachers (Thomas, 1972).

There is, however, no consensus as to the relative effectiveness of different training techniques, and the effects of any procedure are not always predictable. Bailey and Wehr (1971), for example, reported no change in two teacher behaviors when the teacher viewed videotapes of herself teaching and received graphic feedback as to the rates of those behaviors. Since few comparisons of teacher training techniques have been published in the literature of behavior modification, it is difficult to say which procedures might be most effective in teacher training programs.

This study examined the effects of three different techniques for modifying teachers' behavior: instructions plus feedback from the experimenter, video feedback, and a direct intervention procedure. Three behaviors, praise, on-task contacts and off-task contacts, were studied because these behaviors have been reported to affect such student behaviors as studying, following instructions and engaging in appropriate social behaviors in class (Hall, Lund and Jackson, 1968; Madsen, Becker and Thomas, 1968; Schutte and Hopkins, 1970; Thomas, Becker and Armstrong, 1968).

METHOD

Subjects and Setting

The nine subjects ranged in age from 19 to 29 years. Two trainees had one and five years' teaching experience respectively. The other seven subjects had not taught before. One had read some behavior modification literature. Five subjects participated in all phases of the study; four were only available for abbreviated training sequences.

All trainees were volunteers who taught math in one or both of two math classes at the Colorado Springs Community School, an elementary *free* school with an enrollment of 40 students. Each class was one hour long. The first class had 18 students ranging in age from 9 to 14. The 12 children in the other class ranged in age from 5 to 9. The children worked on individualized materials adapted from *Sets and Numbers*, books K through 6 (Suppes, 1962).

The teachers were told that they were participants in a study of the effects of different types of feedback on teacher behavior. The experimenter said she would discuss some procedures with the teachers "at a later time."

Design

The study was a multiple baseline design, both across subjects and across behaviors within subjects. Each training procedure (instructions, videotape and direct intervention) was applied to alter the rate of three teacher behaviors: praise, on-task contacts and off-task contacts. Two behaviors were manipulated in each subject during each condition, while the other behavior was allowed to vary. Experimental conditions began at different times for different subjects.

Definitions and Reliability

A contact was defined as any verbal statement (other than praise) directed to one child. A contact lasted until the teacher moved to another child or until the child spoke. Each contact was scored on-task if directed to a child who was working or raising his hand to request teacher attention, and off-task if directed to a child doing anything else. Praise was defined as any remark such as "that's good," "that's right" or "I like the way you are working."

Videotape recordings were made of each session and scored by the experimenter and an independent observer. Both observers scored each session for the first three weeks. Thereafter, the independent observer scored aperiodically for an average of one out of every five sessions.

The following precautions were taken to avoid experimenter bias in the scoring and to minimize the effect of such variables as the time of the period during which the behavior was sampled. If several subjects were teaching simultaneously, they were videotaped in a different order each day. If a subject was videotaped for more than five minutes, a recording sample was selected by running the tape on a fast forward speed (so that no pre-viewing was possible), stopping the tape after some unspecified time, and recording for five minutes. The independent observer was not informed as to which experimental procedure was in effect.

A frequency count of praise, on-task contacts and off-task contacts (see definitions below) was made by behavioral rather than temporal intervals. An interval began with the teacher's initial contact to a particular child and ended when she spoke to a different child.

The total number of agreements in an interval was the total number of occurrences of each of the three behaviors which was seen by both observers. Disagreements were scored each time one observer scored a behavior not scored by the other observer.

Inter-observer reliability (agreements divided by agreements plus disagreements times 100) on videotapes was 91% with a range from 82% to 100%. Calculated in the same manner, agreement between the experimenter's live scoring and scoring of videotapes ranged from 83% to 90% with a mean of 87%. Agreement between the experimenter and all trainees during the video feedback condition averaged 82% with a range of 71% to 100%.

Stability Criteria

Following baseline, the experimental conditions were applied to each target behavior until the subject's rate of that behavior met two conditions: it showed no trend toward the criterion for three days, and its rate was within 20% of the mean of the previous three days (for behaviors whose rates were 20 responses per minute or higher) or within five responses per minute of the mean (for behaviors whose rates were less than 20 responses per minute). The experimental procedures were applied to the second behavior when the first met the criteria for stability. Because the stability criterion was applied only to the current target behavior, experimental conditions were occasionally changed at a time when the criterion no longer applied to the first target behavior. For example, instructions concerning rate of off-task contacts were given to Subject 1 before session 6. That behavior met the criterion for stability after session 11, so instructions concerning his rate of praise were given before session 12. His praise rate met the criterion for stability after session 17, so a new experimental condition began on session 18 with respect to off-task contacts, even though a downward trend occurred for that behavior during sessions 15, 16 and 17.

BASELINE. On the first day in the classroom, each trainee was instructed to "take over the math group" of three to five children. The trainee was told only that the children were allowed to ask questions about the content of the worksheets (e.g., they could ask the meaning of symbols or how to work particular types of problems). No further instructions were given. Baseline conditions were maintained until observational records indicated that the subject's rates of contact and praise were stable. The trainee received no feedback from the experimenter during this condition.

INSTRUCTIONS AND EXPERIMENTER FEEDBACK. During this condition each subject was given a definition and criterion for the behavior the experimenter was attempting to alter. For example, if the rate of off-task contacts was too high, the trainee was told that it was appropriate to attend only to children who were working or raising their hand to request teacher attention. He was told not to attend to anyone who was not engaging in these behaviors. Consequently, the criterion rate for contacts to on-task children was four per minute, and the criterion rate for statements of praise (contingent on working well or answering correctly) was two per minute.

Throughout this condition graphs of the trainee's defined behaviors were posted in the room, and the ordinate and the abscissa were explained to the trainee. Definitions and criteria for target behaviors were repeated to the subject daily. Manipulation of a second behavior in each subject began after the first met the criteria for stability.

4

VIDEOTAPE FEEDBACK. At the beginning of this condition, the experimenter showed the subject how to record the target behavior. As in the previous condition, the daily instructions to the trainee included the definition and the criterion rate of the behavior. The trainee was required to score five minutes of the videotape of the previous session before teaching. The experimenter scored simultaneously but independently and the trainee and experimenter compared records and scored reliability at the end of the viewing.

Observation of the second behavior began when the first became stable. The trainee continued to score the first behavior when observation of the second one began, but the instructions were given only for the current target behavior.

DIRECT INTERVENTION. Throughout this condition, the experimenter followed the trainee as he taught and continuously recorded his behavior. At the end of each five minute period, the experimenter interrupted the trainee and showed him the recording chart. If the behavior recorded met the criterion, the experimenter praised the trainee and the trainee continued to teach for another five minutes. If the behavior had not met the criterion, the experimenter replaced the trainee and taught for three to five minutes while the trainee recorded. Whether or not the experimenter praised or replaced the trainee during this condition according to whether or not the trainee's behavior met the criterion as prescribed by the procedure was assessed by an independent observer. The agreement between the experimenter and the independent observer as to the appropriateness of each intervention was 100%.

The sequence of conditions was as described for all subjects except Subjects 7 and 9. In order to provide additional information about possible effects of the sequence of experimental conditions, the direct intervention condition was first for these subjects, followed by video feedback. Instructions plus experimenter feedback was omitted from the sequence for these subjects.

RESULTS

The relative effectiveness of the three experimental procedures on praise is shown in Figure 1. Instructions and experimenter feedback had a variable effect on rate of praise. For Subjects 3 and 5, the mean rate of praise increased during this condition, while it decreased for the other two subjects.

The effects of scoring videotapes of oneself teaching were similarly variable. Mean praise rate increased for Subjects 1 and 4 during this condition, decreased for Subject 5, and remained unchanged for Subject 3. During the direct intervention condition, however, the mean rate of praise was above baseline and above the means of all previous

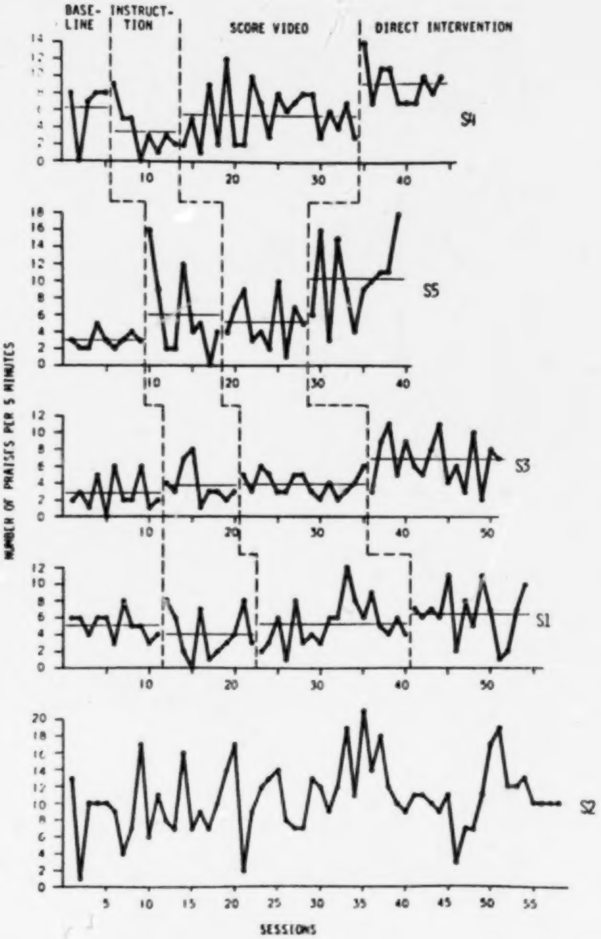


Fig. 1. Rate of praise in four subjects during the experimental conditions.

conditions for all four subjects. The increase in Subject 5's mean rate of praise was small; his mean rate during the direct intervention was 6.4 compared to 5.1 praises per minute during baseline. No systematic change in praise rate occurred in Subject 2, for whom this behavior was never manipulated.

Similar relative effects on the rate of off-task contacts are observed in Figure 2. The effects of instructions plus experimenter feedback are unpredictable. The mean rate of off-task contacts increased in two subjects during this condition and decreased in two others. Recording one's own videotapes reduced the rate of off-task contacts for all subjects, but did not eliminate it. The direct intervention condi-

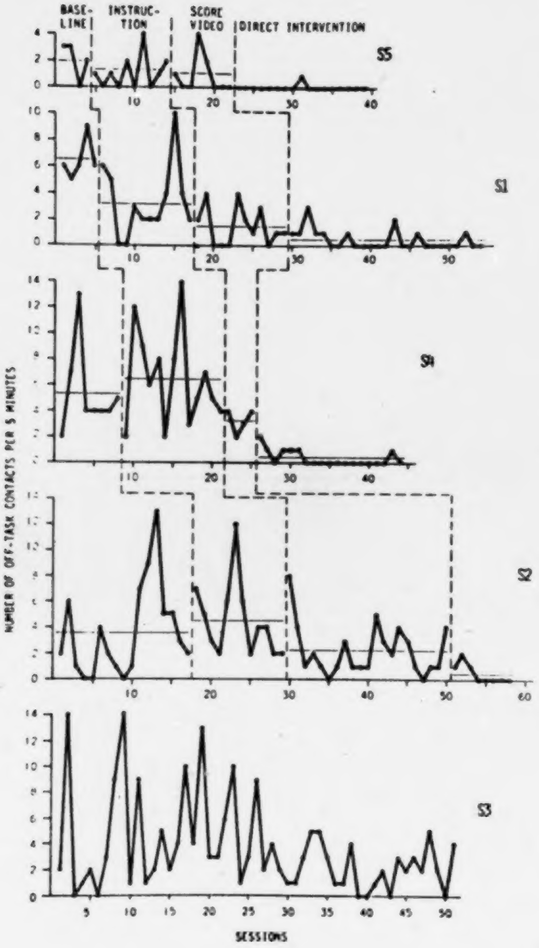


Fig. 2. Rate of off-task contacts in four subjects during all experimental conditions.

tion, however, brought off-task contacts to near zero for every subject. A decreasing trend in Subject 3's rate of off-task contacts was visible, although this behavior was not manipulated. Her rate never stayed at zero for more than two sessions, while the rates of all subjects undergoing the direct intervention procedure remained at zero for five or more consecutive sessions.

Figure 3 shows the rate of praise for Subject 9, for whom the direct intervention condition was the first manipulation, and for Subject 5, who had the same baseline rate of praise. The abscissa is broken because the instructions and experimenter feedback condition was omitted for Subject 9. The highest rate of praise for each subject occurred during the direct intervention condition regardless of this condition's position in the sequence of treatment procedures.

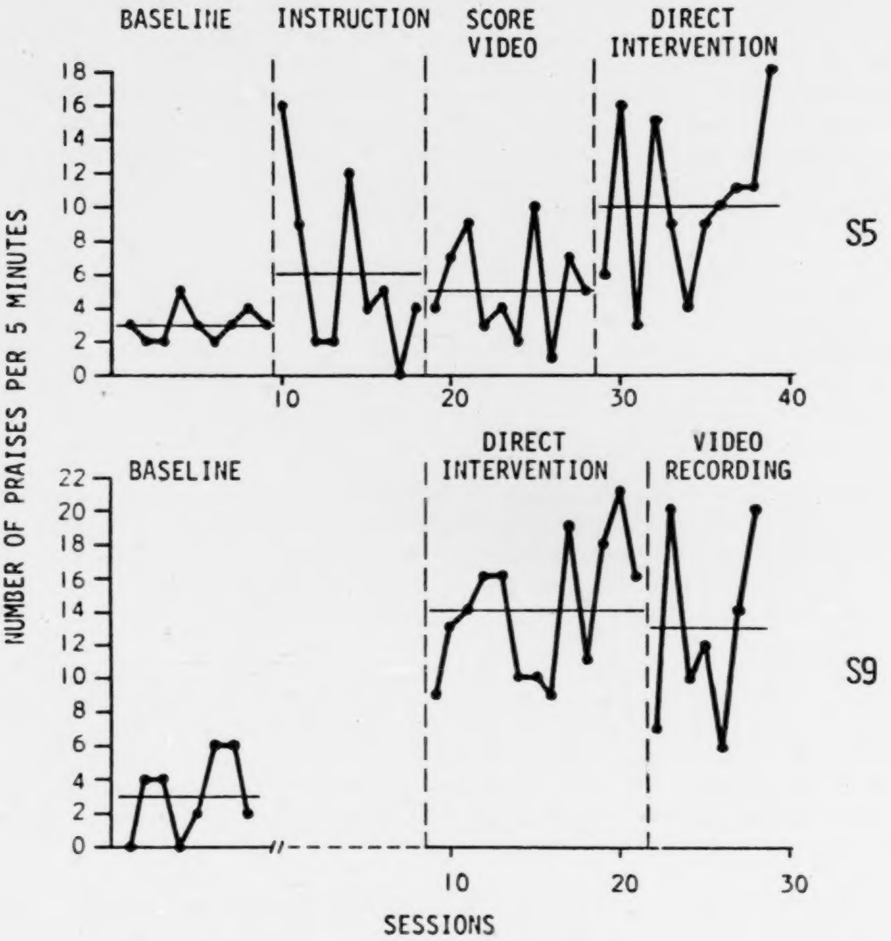


Fig. 3. Rate of praise in two subjects having the same baseline rate when direct intervention was the first and the last experimental condition.

DISCUSSION

The consistent and repeated finding of this investigation was that the direct intervention procedure was most effective in decreasing the rates of undesirable teaching behaviors and increasing those that were desirable. While the procedure was not infallible, it produced marked changes in eight of nine behaviors. The video feedback procedure produced small changes in the appropriate direction in only slightly more than half of the instances where it was applied, and no predictable changes occurred as a function of instructions and daily feedback.

The effects of any of the procedures might have been enhanced if the experimenter had consequences for the subjects' behaviors. All subjects in this study were volunteers who wanted to work in a *free* school (and did not necessarily want to undergo modification of their teaching behaviors). Had the experimenter been in a position to reinforce them socially (as the principals of schools are) or with grades (as do the supervisors of student teachers) or monetarily, any of the procedures might have been more effective.

Direct intervention may have been most effective because it provided the most immediate feedback of the three procedures. Video feedback is necessarily more delayed, and in this case almost twenty-four hours elapsed before the teachers scored their tapes immediately prior to their next teaching session. Thomas (1972) had teachers score tapes soon after teaching and reported systematic changes in a number of behaviors across several subjects.

The differential effectiveness of direct intervention may also be due to the nature of the consequence applied for failure to meet criterion. Comments by the trainees indicated that being replaced was aversive. This may support the suggestion that the consequence of direct intervention is functionally similar to time-out.

Finally, the effectiveness of the direct intervention procedure may have been due to the experimenter's modeling of correct responses for subjects who were replaced in the teaching situation. Because they had to record the experimenter's behavior, the subjects were forced to attend. The effect of modeling might be assessed by comparing the present procedure with one in which subjects are required to leave the room when replaced by the experimenter.

Two considerations concerning praise arose during the study. When subjects' praise rates increased there seemed to be a change in their verbal interactions with the children; they tended to ask more questions and to give fewer explanations. Thomas (1972) reported that praise and instructional prompts were inversely related. Further investigation of the interactions between praise and antecedent behaviors such as prompts seems warranted.

Second, although no explicit instructions were given to praise only children who were working, teachers almost never praised children who were off-task. More than 96% of all praise responses were to children who were working. The problem of changing teachers' praise behavior was not one of teaching a discrimination, as in the case of other verbal contacts, but one of teaching them to increase their rates. The data seem to support the experimenter's impression that teaching the discrimination is easier than teaching teachers to praise frequently.

The results obtained must be qualified in two respects. First, additional examination is needed on the effects of applying these three procedures in differing sequences. One attempt to expand the data on this question was made. The rate of on-task contacts was manipulated using different treatment sequences in Subjects 3 and 7. However, the experimenter increased the criterion on this behavior for Subject 3 prior to the direct intervention procedure. Consequently, the data on possible sequence effects were confounded by the shifting criterion.

A second qualification must be added until more data are available on the durability of the observed effects. In the present study it was not possible to determine the extent to which the improved teaching behaviors would persist in the absence of continued feedback, and whether or not they would generalize to other classroom situations. Although not examined in this study, it is possible that any or all of the experimental procedures may be useful in maintaining appropriate behavior even if they are less effective in supporting the initial acquisition of a teaching behavior.

The results of this study have two broad implications for teacher training. The first is that direct intervention could be a useful on-the-job training procedure for changing teacher behaviors. This sort of training presumably occurs during student teaching, but presently there are few behavioral requirements for student teachers or supervisors. Some student teachers are coached by their supervisors, others clean erasers, and some are abandoned, untrained, to teach the supervisor's class. In order for student teaching to affect teachers' behavior in any predictable fashion, it must include behavioral goals and a procedure to help teachers meet these goals.

Second, if the specific, daily instructions to teachers in this study had little effect on their teaching, the instructions which they receive in methods courses (sometimes preceding the teaching experience by a year or more) cannot be expected to control much classroom behavior.

It seems reasonable to conclude that improvements in teachers' behaviors are best made with on-the-job training that includes procedures designed to teach specific responses.

REFERENCES

Bailey, Jon S. and Wehr, Robert J. The effects of video-tape feedback, modeling and direct remote instructions on teacher behavior: some preliminary findings. Paper presented at the Second Annual Conference on Behavior Analysis in Education, May, 1971, University of Kansas, Lawrence, Kansas.

Broden, Marcia, Bruce, C., Michell, Mary A., Carter, Virginia and Hall, R. V. Effects of teacher attention on attending behavior of two boys at adjacent desks. *Journal of Applied Behavior Analysis*, 1970, 3, 199-203.

Cooper, Margaret L., Thomson, Carolyn L. and Baer, D. M. The experimental modification of teacher attending behavior. *Journal of Applied Behavior Analysis*, 1970, 3, 153-157.

Hall, R. V., Lund, Diane and Jackson, Deloris. Effects of teacher attention on study behavior. *Journal of Applied Behavior Analysis*, 1968, 1, 1-12.

McAllister, L. W., Stachowiak, J. G., Baer, D. M. and Conderman, Linda. The application of operant conditioning techniques in a secondary school classroom. *Journal of Applied Behavior Analysis*, 1969, 2, 277-285.

Madsen, C., Jr., Becker, W. and Thomas, D. Rules, praise and ignoring: elements of elementary classroom control. *Journal of Applied Behavior Analysis*, 1968, 1, 139-150.

Packard, R. G. The control of "classroom attention:" a group contingency for complex behavior. *Journal of Applied Behavior Analysis*, 1970, 3, 13-28.

Schutte, R. C. and Hopkins, B. L. The effects of teacher attention on following instructions in a kindergarten class. *Journal of Applied Behavior Analysis*, 1970, 3, 117-122.

Suppes, Patrick. *Sets and numbers*, Books K through 6. New York: Random House, Inc., 1962.

Thomas, D. R., Becker, W. C. and Armstrong, Marianne. Production and elimination of disruptive classroom behavior by systematically varying teacher's behavior. *Journal of Applied Behavior Analysis*, 1968, 1, 35-45.

Thomas, Don Roy. Self-monitoring as a technique for modifying teaching behaviors. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, 1972.

Wasik, Barbara H., Senn, Kathryn, Welch, Roberta H. and Cooper, Barbara A. Behavior modification with culturally deprived school children: two case studies. *Journal of Applied Behavior Analysis*, 1969, 2, 181-194.

3 STUDENT STUDIES

patricia m. pierson

INTRODUCTION

The three studies reported here were done by a student, with a student, and for a student. They were conducted during my first year as a consulting teacher in training at the University of Vermont. Prior to the beginning of the first study, my only training had been a six-week summer school course in Behavior Analysis. Twelve years before, I had taught in an elementary classroom for three years but the intervening time had been spent in the diaper pail and behind the ironing board.

The studies were completed as part of a course requirement. The first study was done by Miss Phyllis Marcell, a student in a graduate course in Behavior Analysis. The other two studies were completed as part of a certification workshop that I planned and taught. The teachers were Miss Barbara Butler and Mrs. Constance Miles. Constantly on hand during all three studies with advice and encouragement were my advisor, Dr. Edward Hanley, and Mrs. Betty Jane Lates, consulting teacher for the Burlington School district.

EXPERIMENT I

Subject and Classroom

Fred was a nine-year-old boy who had been in an Educable Retarded classroom for the past two years. He scored 71 on the WISC test, had a reading level of 1.3, a spelling level of 2.2, and an arithmetic level of 2.6, in Spring, 1970. Twelve children were in the class, ranging in ages from seven to ten. The classroom was part of a regular elementary school.

Behavior

Fred did not complete his work assignments although the teacher knew he was capable of doing so. He tired easily, was often sleepy and was frequently absent due to illness. The behavior desired was defined as completion of arithmetic work during a 20-minute period with 80% accuracy.

Measurement and Reliability Procedures

The teacher recorded daily the percent of problems completed and the percent of problems correct. Reliability was obtained by an observer checking the teacher's computations weekly. The student's work was done in a notebook, and consisted of two digit arithmetic problems which were written by the teacher. The student wrote his answers in the book. Due to the simplicity of the assignment and the fact that it was not corrected by the child after the teacher marked it, the reliability was always 100%.

Teaching/Learning Methods and Materials

Problems given were two and three place addition and subtraction without regrouping. Twenty problems were assigned each day.

Baseline 1

The teacher assigned work and recorded daily the percent done and the percent correct.

Contingency I

If Fred finished ten problems with 80% correct, he received one chip; he received another for the next ten done with the same accuracy. Two chips were used to buy food to be eaten at snack time. No chips could be accumulated toward the next day.

Scientific Verification

The teacher stopped giving Fred the chips; hence, he could not buy a snack.

Contingency II

The teacher resumed giving Fred the chips and he was again able to buy a snack.

Postcheck

Four months later, the teacher noted that Fred continued to finish his work and scored 80% or above.

RESULTS

Figure 1 shows the number completed of assigned arithmetic problems. During the baseline period, Fred completed an average of twelve problems with a range from nine to fifteen. During the first contingency his average rose to eighteen problems with a range of ten to twenty. During scientific verification, when he no longer received tokens, his average dropped to ten with a range of seven to twelve. When tokens were reinstated in contingency II, his average number done rose to nineteen with a range of six to twenty. Postchecks showed he was continuing the good word.

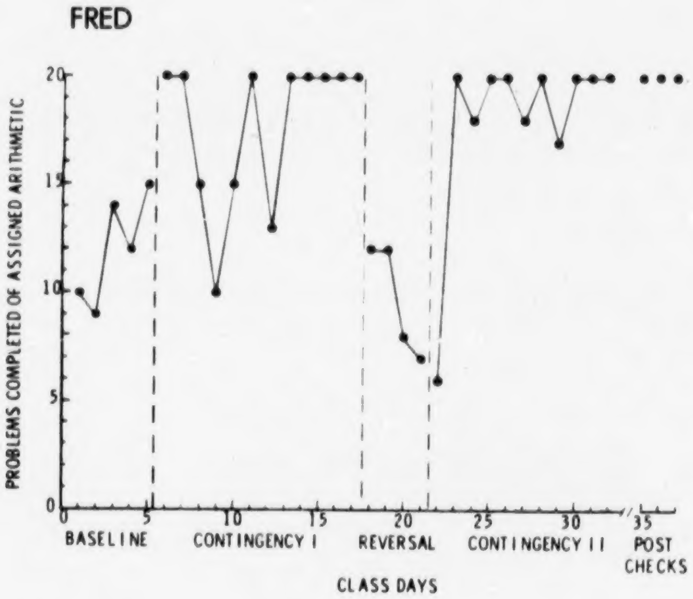


Fig. 1

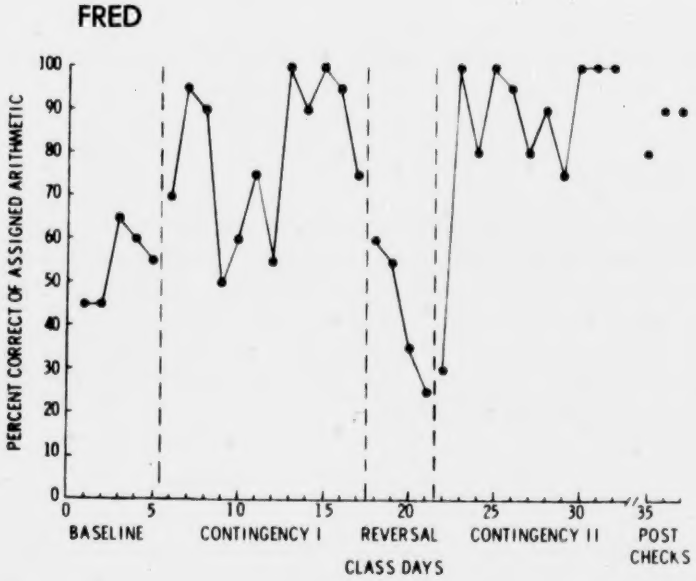


Fig. 2

Figure 2 shows that the percent correct generally followed the same pattern as the number completed. During both contingencies Fred was 80% or more accurate fifteen out of twenty-three times.

DISCUSSION

The use of tokens with snacks as backup reinforcers was found to be an effective system for Fred. The teacher then decided to use these procedures on a daily schedule. Fred got one token for his morning's work being completed, one for reading work completed and a third for completion of afternoon work. He was allowed to exchange tokens for a snack at the end of the day. The teacher reported he had earned his snack every day except one when she was absent and not there to administer the tokens.

EXPERIMENT II

This study was done in Spring, 1970, as part of a certification workshop.

Subject and Classroom

Patty was a seven-year-old girl in a first grade class from a middle income area. She had had problems in school kindergarten and was reported to have spent most of that year hiding under a table. She had been referred to a family counseling service where no specific reason was found for her behavior. This was her second year in first grade, her first year in this school. She was tested on the Wechsler Test in June, 1970, and scored 80-90.

Behavior

Patty's social behavior was greatly improved in school. She took part in class activities and was well liked by her classmates. She was unable, however, to follow directions or complete her written work so that, even though she was a good reader, her grades were poor. Her behavior at home was erratic and her parents reported that she was difficult to handle.

Reliability Procedures

Patty's work was corrected immediately upon completion, and a token card was given. During Contingency II, when the card could be exchanged for gifts in 40%, 60%, 80% or 100% categories, the cards were made out to the nearest of those percents, and if the score was in the middle, the higher category was granted. An observer checked Patty's workbook weekly. Because responses were highly objective and the teacher was very precise, there was always 100% reliability.

Teaching/Learning Methods and Materials

Patty was reading in the Scott Foresman book, *Level 1*. After the session with the reading group, the teacher would explain the workbook pages and the children would do them at their seats.

Baseline 1

During the baseline, the teacher kept a record of the percent done and the percent correct of Patty's reading workbook assignment.

Contingency I

When Patty completed all of her reading workbook assignment with at least 75% correct, she got a happy face card to take home. Her mother would give her five cents for each card. After eleven days it was clear that this contingency was not strong enough to modify Patty's behavior.

Contingency II

Patty's mother took her shopping and she chose toys and candy of varying worth which were kept on a shelf at home. Examples were: packages of gum; one stick in the 40% category, two sticks in the 60% category, etc., play make-up set; powder in the 40% range, lipstick in the 60% range, etc. The 100% category prize was a "Dawn" doll.

Patty had to get 100% of her work completed, and was given a happy face card for 40%, 60%, 80% or 100% correct. These were exchanged at home for the gifts on the shelf in the proper percent category.

Reversal

The teacher was unexpectantly absent and Patty was unable to get her card to take home.

Contingency III

Patty continued earning cards when 100% of her workbook assignment was completed and was given a happy face card in the percent category she had earned. These cards continued to be exchanged daily for gifts at home, in the appropriate percent category.

RESULTS

Figure 3 shows percent of work done. During baseline, Patty varied between 0% and 100%, with an average of 50% completed. Contingency I ranged from 0% to 100% with 49% average. During Contingency II, Patty completed both assignments, but during Reversal she completed only an average of 12% of her assignments. Percent completions varied between 50 and 100 in Contingency III with an average of 97%.

The data for percent of work done correctly is shown in Figure 4. During baseline, Patty varied between 0% and 70% correct with an average of 35% correct. During Contingency I the range was 0% to 90% correct with an average of 37% correct. Contingency II showed a range of 54% to 75% with an average of 64%. During the Reversal period, scores were 10% and 47% correct.

In Contingency III, the range was 12% to 100% correct with an average of 70% correct.

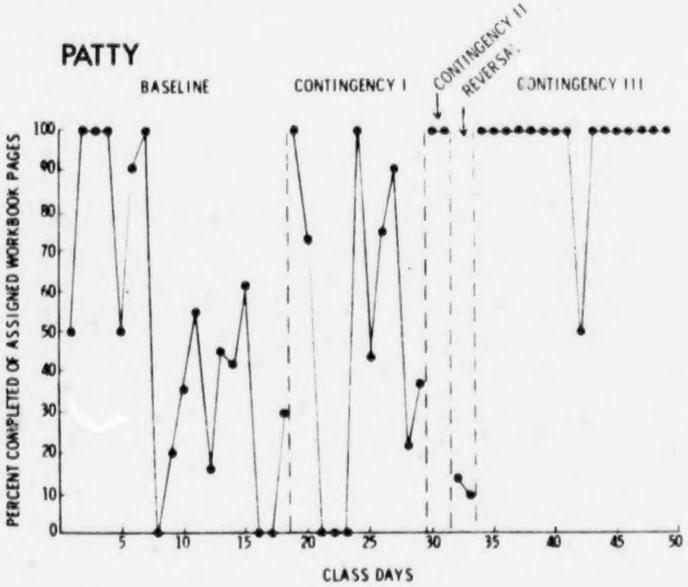


Fig. 3.

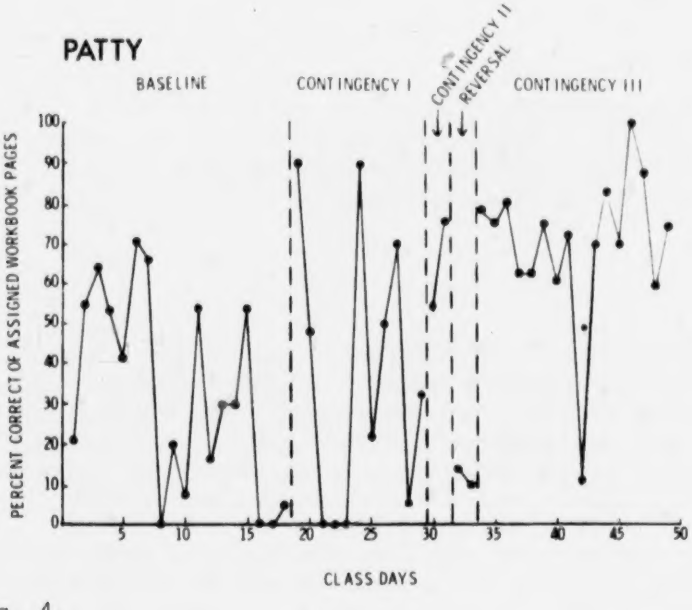


Fig. 4.

CONCLUSIONS

This study shows a dramatic increase in the amount of work Patty completed during Contingencies II and III, and a definite increase in accuracy. The unplanned Reversal provided a verification of the reinforcing effects of the cards. When the teacher returned during Contingency III, Patty immediately resumed her high percent of work completed and continued to earn a card for at least 60% correct every day but one. The parents reported a great change in Patty's behavior at home. She was more stable, completed household tasks, and showed more initiative around the home. After Patty had earned her 100% correct gift, the "store" was replenished and the cards were made contingent upon doing the entire morning's class work. Patty did not keep up her former level of work done and accuracy, but the parents reported that her home behavior was still good and the procedures are being continued.

EXPERIMENT III

This study was completed during Spring, 1970, as part of a certification workshop.

Subject and Classroom

Connie was a seven-year-old girl in the first grade. The preceding Fall, Connie had tested 36 (low-medium) on the Metropolitan Readiness Test. This study was carried out in the learning center where Connie's class met daily.

Behavior

Connie attempted little of the assigned work during a 30-minute learning center period. When she hurried through the work at the end of the period, it was inaccurate. Connie's printing was *poor*. Her teacher asked her to copy a model letter, upper case on one side of the paper, lower case on the other as many times as possible in a 15-minute period. Each day one letter was given, starting with *A* and going in alphabetical order. The first five letters given during the baseline were repeated again when contingency was started. The three letters assigned during the Reversal period were also repeated, at Connie's request.

Measurement and Reliability Procedures

The teacher counted the number of correctly formed letters, on the basis of the overlay. An observer checked the papers once a week, also using the overlay. Because of the thicker overlay letters, it was very simple to judge whether or not the letter should be counted as accurately formed; thus, agreement was always 100%.

Baseline 1

The teacher gave Connie a pencil, an eraser, and lined paper with one model upper case letter on one side and one model lower case letter on the other. At the end of a 15-minute period, the teacher collected the paper.

Contingency I

The teacher made a special loose leaf notebook for Connie. When Connie accurately formed at least ten letters during the 15-minute period, she was allowed to put the page in the book and take it home. The teacher also went to Connie at 3-minute intervals and checked her work for accuracy.

Scientific Verification

The teacher kept Connie's notebook. Furthermore, she did not make checks for accuracy at 3-minute intervals.

Contingency II

For the first four days, the teacher reinstated the 3-minute accuracy checks. The notebook was then returned on the fifth day.

Thinning

The time period was shortened to 7 1/2 minutes, and the teacher made one accuracy check in the middle of the period.

RESULTS

Figure 5 shows that during baseline the number of accurately formed letters ranged 0 to 3, with an average of 1. During Contingency I, there was a range of 8 to 21 correct letters, with an average of 14 correct. During Scientific Verification, correct letters ranged from 1 to 18, with an average of 4. In Contingency II, the range was 7 to 25, with an average of 15. When the Thinning process was instituted, the range was 10 to 19, with an average of 13 correct letters.

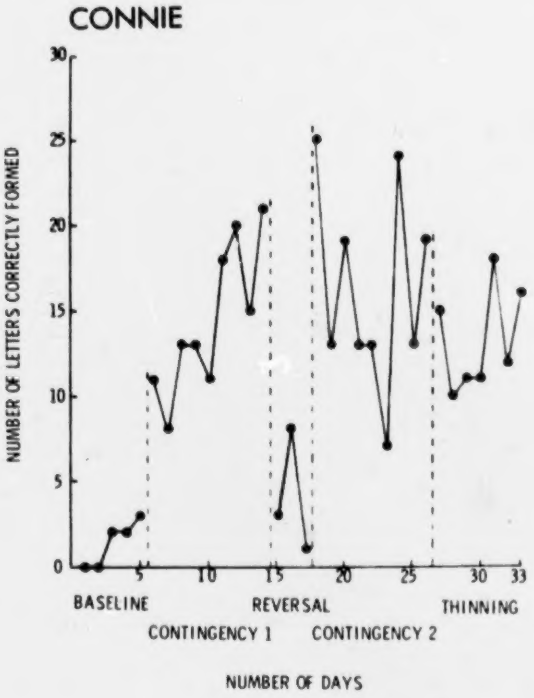


Fig. 5.

DISCUSSION

Connie's rate of correct letter formation increased during both contingency periods. The 3-minute checks by the teacher, when reinstated alone during the second contingency for the first four days, showed that this was as effective a reinforcer as the notebook. The rate stayed almost exactly the same when the notebook was returned. During the Thinning procedure when Connie was allowed half as much time to work and the teacher made one check, she still maintained an average of 13 letters.

The teacher reported that Connie took great interest in her writing work; she asked to do over the letters that were assigned during the Reversal period, and she was able to tell the teacher which would be the correctly formed letters before they were checked with the overlay.

PART 6:

PROGRAMS & TECHNIQUES IN BEHAVIOR ANALYSIS:

potpoURRi

THE EXPERIMENTAL ANALYSIS OF A HOMEWARD EDUCATIONAL PROGRAM WITH A PRE-DELINQUENT JUVENILE

martin j. pollack
beth sulzer-azaroff
reed williams[1]

Education related behaviors have been examined by manipulating tests and homework using varying schedules of presentation (Malott and Svinicki, 1969; Bostow, 1970; Bostow, Mawhinney, Blumenfeld and Laws, 1971). These studies have concentrated on normal populations. Results generally indicate greater consistency and higher rates of behavior for daily schedules than for intermittent schedules. Additional research has established the efficacy of using behavioral objectives in instructional settings (Mager and Clark, 1963; Miles, Kibler and Pettigrew, 1967). In the present research, the independent variable was the manipulation of a schedule of contingencies in the form of assignments. Daily and weekly assignments were contrasted and performance was measured in terms of number of assignments completed to determine whether homework production would be differentially affected by the schedule used. A token system using points and home reinforcement procedures was added as part of a program designed to bring the student under control of the contingencies for academic work. The program was applied here to a *pre-delinquent* high school youth who was out of school at the time of the study.

METHOD

Subject

The student was a sixteen-year-old male classified by the high school as *pre-delinquent*. Behavioral characteristics included a sparse verbal repertoire, little activity and, especially, excessive sleeping. No medical justification for this condition was revealed by an examination. The home environment was a compact living space for both parents and six siblings, but showed no evidence of economic deprivation. The student had a history of declining attendance over the past two years and a "D" academic average despite normal range intelligence test scores. Deficiencies were particularly severe in spelling and mathematics. The student had been expelled for truancy at the start of the study.

---

[1] Grateful acknowledgment is extended to David Miles, John Deichmann and Jerry and Jean Ulman for their helpful suggestions and comments.

Experimental Setting and Materials

The experimenter visited the student at home daily by agreement with the student and the student's parents. Programmed texts in spelling and mathematics (Hook, 1967; Hauck, Moore and Smith, 1966) were used together with written assignments which included behavioral requirements and due dates. Problems and sentences using spelling words were the operations required. Tokens in the form of points were provided contingent upon completion of assignments on time for two academic subjects and were exchangeable for back-up reinforcers provided by parents and the experimenter. A list of reinforcer prices (see Table 1) and records indicating academic progress in terms of assignments completed were available to all parties.

TABLE 1

LIST OF REINFORCERS AND PRICES

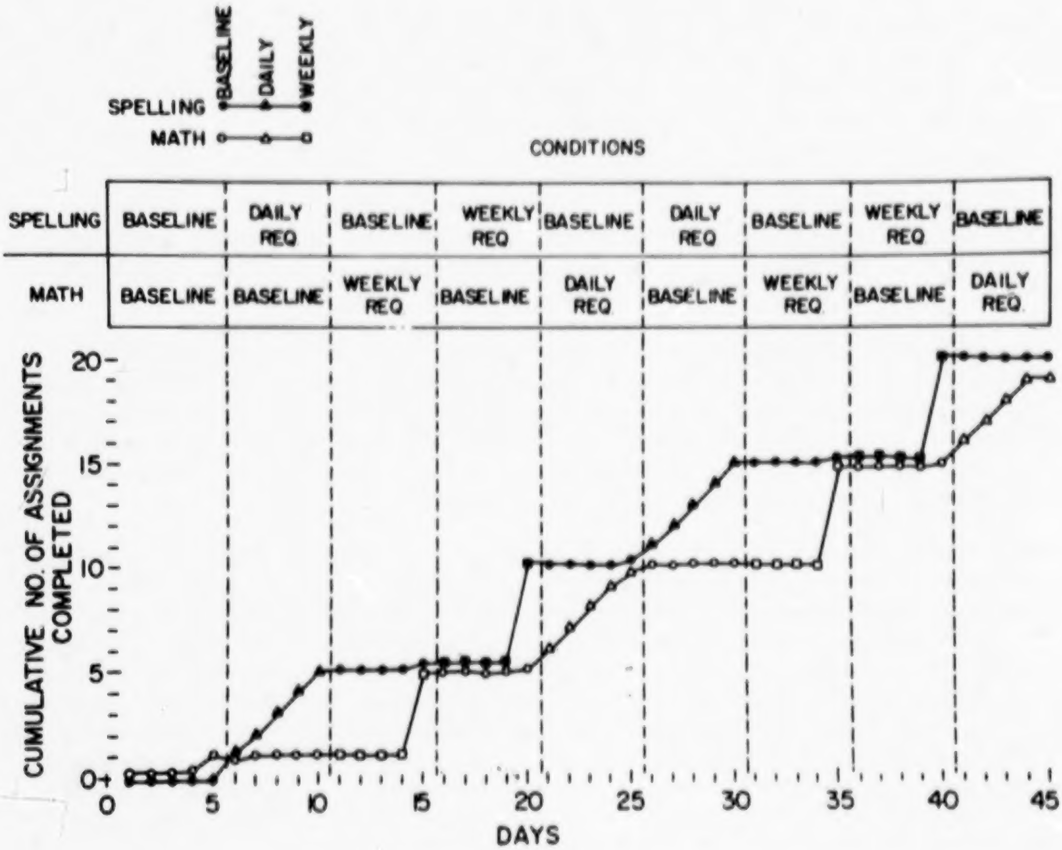
REINFORCER	POINTS REQUIRED	CONDITIONS
● USE OF CAR	*90*	*Time Specified*
● MOVIE	*100*	
● ATHLETIC EVENT	*100*	
● ACCESS TO MECHANIC	*90*	*Per Hour*
● BOATING	*90*	*Per Hour*

Procedure

Diagnostic testing was done with sample materials for each academic category prior to the start of the program. The experimenter was thus assured that the student could perform the tasks involved and was able to function at the level demanded by the program. The student was started at the beginning of each program based on scores of 38.7% correct for spelling and 60% correct for mathematics. The program employed a token (point) reinforcement system in which 100 points were delivered to the student each week contingent upon punctual completion of all assignments, either daily or weekly, for that week. Daily sessions included recording of work completed, testing over assignments when completed, requests for token exchanges, additions to the back-up reinforcer list and answering questions. Daily records of completed assignments were kept and cumulative graphs were drawn as the program progressed. The terminal performance was specified as the completion of all requirements for each assignment.

The experimental design consisted of two counter-balanced reversal designs in spelling and mathematics (Baer, Wolf and Risley, 1968). Scheduling of conditions for the two subjects was varied according to the sequence indicated in Figure 1.

The program was arranged so that during the same week in which the baseline condition was in effect for one subject, one of the two experimental conditions was in effect for the other subject. The exper-



CONDITIONS

SPELLING	BASELINE	DAILY REQ	BASELINE	WEEKLY REQ	BASELINE	DAILY REQ	BASELINE	WEEKLY REQ	BASELINE
MATH	BASELINE	BASELINE	WEEKLY REQ	BASELINE	DAILY REQ	BASELINE	WEEKLY REQ	BASELINE	DAILY REQ

Fig. 1. Cumulative graph of assignments completed across conditions for each academic subject.

imental conditions alternated from subject to subject (spelling to mathematics to spelling....), and within each subject between daily assignments and weekly assignments with succeeding weeks. Baseline conditions were interspersed between each experimental condition for each subject.

Baseline (No Assignment). Reinforcement was not available for any work completed. Assignments were not given, so the student could work in the materials as desired, or do no work whatever. Daily records were kept of any work completed.

Daily Assignments. Reinforcement occurred only if each of five assignments for the week were completed as specified. Verification of assignment completion was made via daily data records. Delivery of points took place at the end of the five day week. One hundred points were given for completion of all five daily assignments on time. Partial or non-completion of any one of the five assignments resulted in no points, regardless of other assignments completed.

Weekly Assignments. Reinforcement occurred only if the large single assignment was completed entirely. Delivery of points took place at the end of the five day week on the same day the assignment was due. One hundred points were given for completion of the single assignment. Partial or non-completion of the assignment resulted in no points.

The three conditions and the reinforcement system were explained to the student at the outset. For each of the subjects, assignments and completion dates were specified orally and in writing on behavior unit forms. The student was asked to present all math or spelling work completed (at the start of each daily meeting, regardless of the schedule in effect or the academic material assigned).

The number of completed items required for assignments given from week to week was kept constant insofar as possible. This allowed for a comparison of performance in different weeks and under different conditions. Weekly assignments contained approximately 100 items. In order to maintain comparability between daily and weekly assignments, daily assignments were constructed with approximately 20 items each. Deviations from these figures were due to attempts to maintain continuity in the materials. For example, an assignment which consisted of precisely 20 items but which fell two or three items short of completing a unit, or extended into the next unit by a similar number, were altered appropriately in an upward or downward direction.

Failure to complete assignments had no effect on the choice of materials for succeeding weeks. Recycling of materials was rejected since such a practice might work counter to the contingency which required completion of assignments by a specified date. Past materials were, however, available to the student as a potentially useful aid in completing assignments for which contingent reinforcement did exist. The student could purchase reinforcers as long as he had earned enough points to pay for them. Those reinforcers which took the form of events or priveleges were arranged as expediently as possible for a mutually agreeable time.

At the completion of the program, an unannounced retention test consisting of representative items from each of the tests over individual assignments was administered for each subject.

Results

The results show that completion of assignments was controlled in a highly predictable fashion. Daily assignments produced steady daily performance while weekly assignments produced a pattern that accelerated immediately before the due date. Over a nine-week period, all four weekly and nineteen of twenty daily assignments were completed as specified. Performance on only two of the twenty-three tests fell below 70% correct. The mean across conditions was 85.8% correct for spelling and 77.2% correct for mathematics. The mean correct for all daily spelling and mathematics tests was 80%, while the mean for all weekly tests was 90% correct.

Retention test scores were 62.5% correct for spelling and 82.5% correct for mathematics. These scores represent a positive change of 23.5% and 22.5%, respectively, over entering test scores. Error rates were not noticeably different for daily and weekly schedules.

Discussion

The data present clear evidence that control and prediction of this student's homework behavior could be achieved by manipulating tests and the time requirements placed on assignments. The data lend further support to the contention that academic performance can be brought under schedule control.

An analysis of the motivating factors in the program is appropriate. The student accumulated a total of 700 points but did not attempt to exchange them for back-up reinforcers until the seventh week when he purchased mechanics training. This occurred despite the fact that the student could choose reinforcers at any time. This suggests that the back-up reinforcers might not have been crucial for the student. Some evidence is available to suggest that points alone can temporarily serve as reinforcers (Sulzer, Hunt, Ashby, Krams and Koniarski, 1971). Another possible source of reinforcement was in the form of attention from the experimenter via his daily contact with the student throughout the program.

The presence of tests consistently following each assignment is a variable worthy of consideration. Since the student scored on the average above 80% correct on the tests, a performance considerably better than recent school experiences, the tests may have functioned as a reinforcing event following the completion of the assignments. Individual or cumulative effects of these potential reinforcers may have contributed to the control achieved.

The data also provide evidence to show that a large amount of materials could be completed with assignments with longer time requirements. There was no difference in total production with daily and weekly assignments with the exception of the last day. Given that future research does not indicate differential retention associated with the two conditions, it would appear that the weekly assignment is more desirable if positively accelerated patterns of responding are acceptable.

The effectiveness of the procedure for applied settings is emphasized by its success with a student whose academic history includes a negligible amount of completed homework. The desired patterns of responding were obtained by using the appropriate schedule. During all four weeks of daily assignments, the student completed homework for four successive days in the absence of any reinforcement (tokens) built into the system. Only on the fifth day could his efforts be rewarded. The apparent willingness to wait for the payoff is particularly significant for this individual since he has progressed from no measurable performance to specified performance under a schedule as lean as a weekly interval schedule.

Attempts to generalize these results through replication with a group design would be a logical next step. The length of the interval could be explored as well to determine the one most efficient for a program. Variability in retention as a function of changes in length of the interval would be an important consideration in choosing the schedule most suitable for a particular setting.

REFERENCES

Baer, D. M., Wolf, M. M. and Risley, T. R. Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1968, 1, 91-97.

Bostow, D. E. The effect of pop quizzing vs. daily quizzing on undergraduate students' studying behavior. Unpublished doctoral dissertation, Southern Illinois University, 1970.

Bostow, D. E., Mawhinney, V. T., Blumenfeld, G. J. and Laws, D. R. An analysis of student studying behavior as a function of two testing schedules. Paper presented at the annual meeting of the American Educational Research Association, New York, March, 1971.

Hauck, W., Moore, J. W. and Smith, W. S. *Decimals and percentage: a programmed unit in modern mathematics*. New York: McGraw-Hill, 1966.

Hook, J. N. *Spelling 1500: a program*. New York: Harcourt, Brace and World, 1967.

Mager, R. F. and Clark, C. Explorations in student-controlled instruction. *Psychological Reports*, 1963, 13, 71-76.

Malott, R. W. and Svinicki, J. G. Contingency management in an introductory psychology course for one thousand students. *The Psychological Record*, 1969, 19, 545-556.

Miles, D. T., Kibler, R. J. and Pettigrew, L. E. The effects of study questions on college students' test performances. *Psychology in the Schools*, 1967, 4, 25-26.

Sulzer, B., Hunt, S., Ashby, E., Krams, M. and Koniarski, C. Increasing rate and percentage correct in reading and spelling by means of a token system with a fifth grade class of slow readers. Paper presented at the Second Annual Conference on Behavior Analysis in Education, Lawrence, Kansas, May, 1971.

THE AUTO-PEDESTRIAN SAFETY PROJECT

donald a. jackson
william j. mayville
joseph b. cowart, jr.

The analysis of five years of police accident data for Salt Lake City showed that there had been a yearly increase in the number of traffic accidents involving automobiles and pedestrians. The police data also showed that more children between the ages of five and nine (about twelve percent of the population) were involved in traffic accidents each year than any other comparable age group. Although these data do not reflect the number of people in the different age groups that were susceptible to accidents by their mere presence on sidewalks or streets, this low age group was chosen for pedestrian safety training because of the apparent high incidence of accidents in this group and because of the great interest and support for pedestrian safety in children.

The notion of changing the pedestrian behavior of elementary aged children was certainly not new. Most Salt Lake schools had some sort of program aimed at making children become good pedestrians. However, the effectiveness of these attempts had never been shown. It was hypothesized that the apparent lack of effectiveness of these programs was due to the absence of any direct modification procedures. It seemed unlikely that a child would learn to stop and check traffic at a street corner simply because of verbal warnings or instructions given in a classroom, especially in the absence of any reinforcement contingencies. The safety programs conducted by elementary schools (and other groups) in Salt Lake City did not include any direct manipulations of reinforcers related to pedestrian behavior itself. The purpose of this study, then, was to demonstrate a procedure by which the pedestrian behaviors of elementary-aged children could be modified by school personnel or volunteers and to show that the changes in pedestrian behaviors observed were functionally related to the procedures used.

METHOD

Subjects

The subjects were male and female elementary school students who attended a summer program at Bonneville School in Salt Lake City. More specifically, those students who left the school between 9:45 AM and 10:10 AM, 10:45 AM and 11:10 AM or 11:45 AM and 12:10 PM, and crossed at or between corners A, B and C in Figure 1 participated in the safety project.

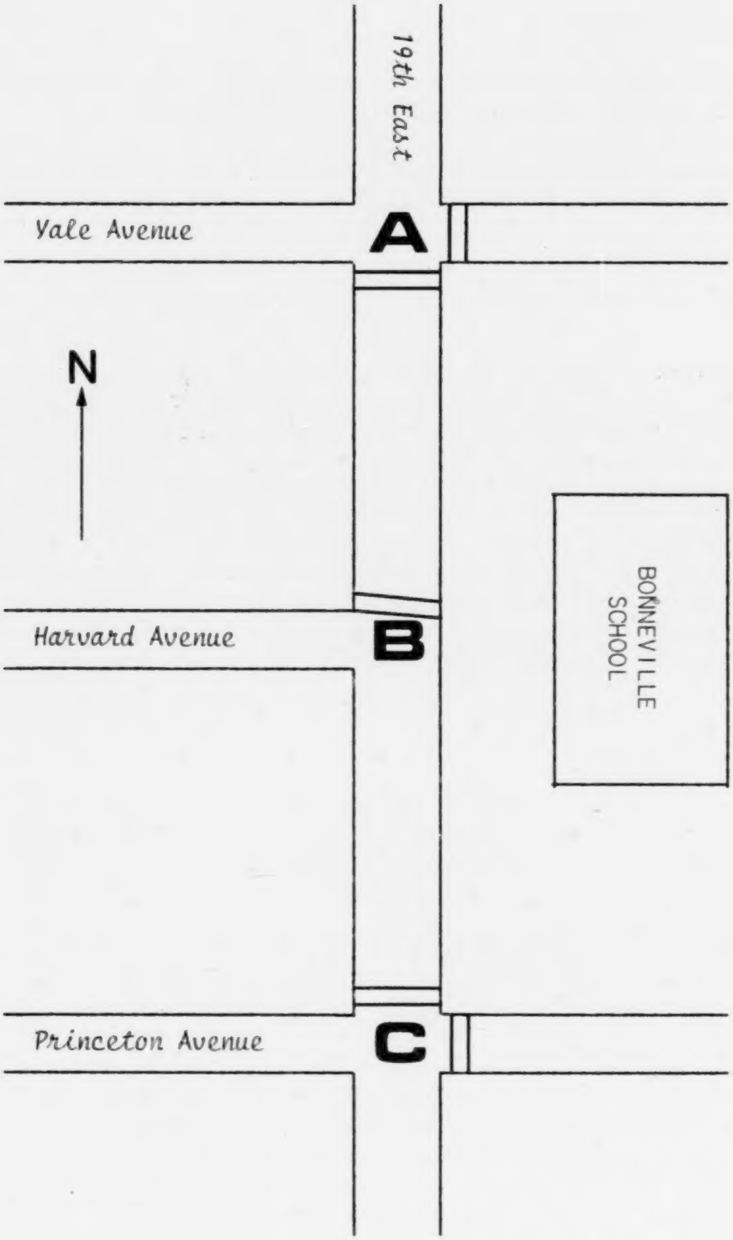

19th East
Yale Avenue
A
N
Harvard Avenue
B
BONNEVILLE SCHOOL
Princeton Avenue
C

Fig. 1. Intersections in Experimental Area at Bonneville School.

Apparatus and Situation

The project required the following equipment: ten single event hand counters, assorted pads and pencils, and clipboards.

The three intersections nearest the school (see Figure 1) and the street area between these intersections were designated as the experimental area.

Personnel

Eight female high school students functioned as data collectors and behavior modifiers. The girls were trained and their subsequent project behavior was monitored and modified by three project coordinators from the Bureau of Educational Research at the University of Utah. One teacher from the Bonneville School teaching staff (who was also the summer school coordinator) assisted in scheduling assemblies and facilitating communication between school personnel and the project coordinators. A University of Utah work-study student directly supervised the high school students in many of their duties and assumed the added responsibility of assisting in the collection and compilation of the data. (If this were not an initial, research study, fewer personnel would have been needed.)

Procedure

OBSERVATION. Prior to the formal collection of data, motorist and pedestrian behaviors were observed. As a consequence of this observation period, pedestrian street crossing behavior was designated as the dependent measure. Appropriate street crossing behavior was defined as:

> A pedestrian approaching an intersection on foot was to break stride and come to a complete stop at the curb of the intersection, make discreet observing responses in all directions of potential traffic flow for a period of at least three seconds, and then proceed across the street within the confines of the pedestrian crosswalk. In those instances where an adequate view of potential traffic flow was obstructed by parked cars the pedestrian was to proceed (after a three second observing response at the curb) slowly into the street until a safe position with an unobstructed view of traffic had been reached close to the traffic side of the parked cars, then pause for another three seconds and make discreet observing responses before crossing the street. In addition to the above requirements, bicycle riders were to dismount at the intersection curb and walk their bicycles across the street.

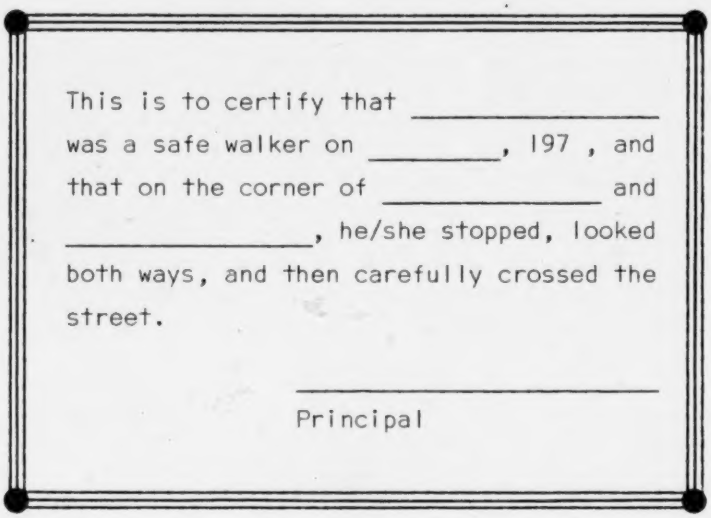
BASELINE. During the baseline period, data were gathered at intersections A, B, and C. The percent appropriate crossings occurring at the three intersections was tabulated separately for each of the three time designations. Only pedestrian crossings from the block that the school was on were counted. Data collectors were positioned so as to have an unobstructed view of the five crosswalks shown in Figure 1 and of the street area in between intersections A and C on 19th East. Any student leaving the school block within fifteen feet of the corners of 19th East and Yale or 19th East and Princeton were observed and counted also. The total number of street crossings and the total number of appropriate crossings were recorded on hand counters between 9:45 AM and 10:10 AM, between 10:45 AM and 11:10 AM, and between 11:45 AM and 12:10 PM.

Reliability checks were taken during each of the experimental conditions by having a second observer, usually one of the project coordinators, record crossings independently of the observer assigned to that intersection.

MODIFICATION. Each of the three time periods (9:45 AM - 10:10 AM; 10:45 AM - 11:10 AM; and 11:45 AM - 12:10 PM) was designated as a separate experimental condition. In accordance with a multiple baseline procedure (Baer, Wolf and Risley, 1968) modification procedures were begun separately for each of the three time periods. On the day that modification was to start for the children leaving between 9:45 AM and 10:10 AM, an assembly was held for these children. The assembly program included a short lecture by a project coordinator, a brief question and answer period, a demonstration of appropriate street crossing behaviors, and a role playing period during which several students were chosen from the audience to demonstrate correct street-crossing behavior. The assembly programs lasted about 20-30 minutes.

Starting with the day of the assembly, the children leaving the school between 9:45 AM and 10:10 AM were reinforced for crossing the street in the manner demonstrated in the assembly, and as defined previously. The person who dispensed reinforcers would greet the child as he completed his crossing by saying, "That was a good job of crossing the street;" or something similar, and would then give the child a piece of candy, a smile, and a *Good Pedestrian Citation* (see Figure 2). The same procedures were used at all three corners. An observer recorded the total number of crossings and the number of appropriate crossings. Baseline recording continued during the 10:45 AM - 11:10 AM and 11:45 AM - 12:10 PM periods.

Two days after the start of the modification procedures for the 9:45 AM - 10:10 AM group, the same assembly was held for all those children who



This is to certify that ________________
was a safe walker on ________, 197 , and
that on the corner of ______________ and
________________, he/she stopped, looked
both ways, and then carefully crossed the
street.

____
Principal

Fig. 2. Example of "Good Pedestrian Citation."

left the school between 10:45 AM and 11:10 AM. Reinforcement procedures for appropriate street crossings were started on all three corners just as they had two days earlier for the students leaving at the earlier time. Baseline recording continued for the 11:45 AM - 12:10 PM period.

Three days after the start of the modification procedure for the 10:45 AM - 11:10 AM time period, the same assembly was held and the reinforcement procedures were implemented for the children leaving the school between 11:45 AM and 12:10 PM. At this time, and continuing until the end of the project, the reinforcement procedures were in effect on all three corners during all three time periods.

RESULTS

Baseline

Figure 3 shows the percent of appropriate street crossings during baseline for all three time periods. The percent of appropriate crossings was low and stable for all three groups.

Modification

The effect of the modification procedures is also shown in Figure 3. The percent of appropriate crossings during the 9:45 AM - 10:10 AM period increased greatly in session seven, the first day of treatment.

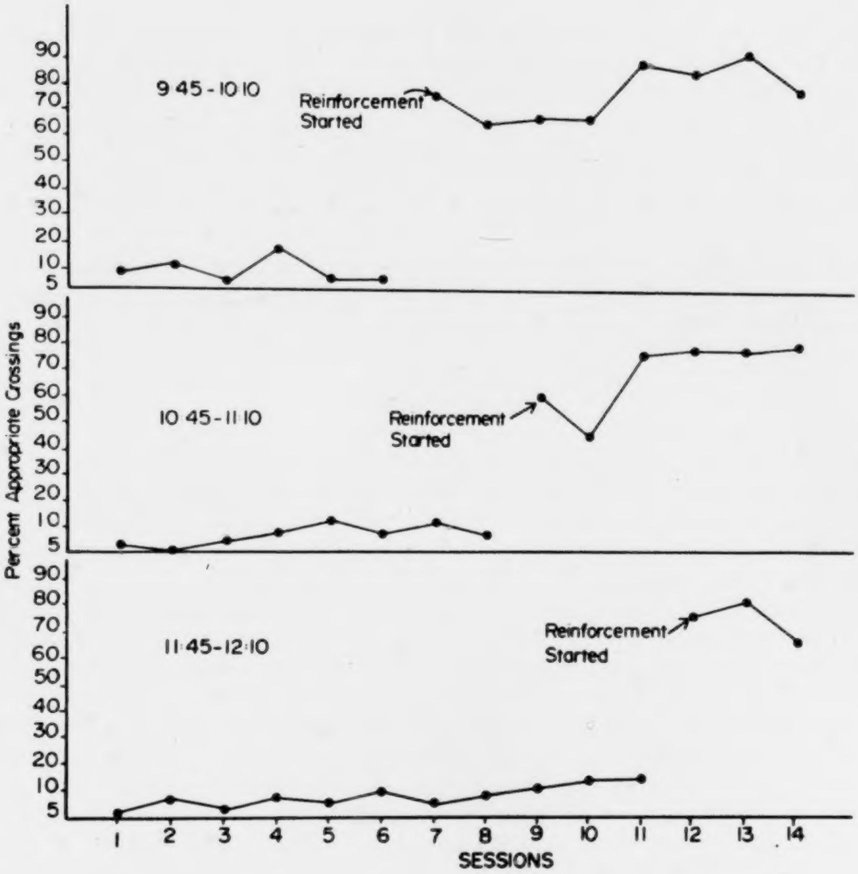


Fig. 3. Percent of appropriate crossings as a function of sessions.

The rate of appropriate crossings remained high, never falling below 63% for the remainder of the project. Similarly, when the modification procedures were instituted for the 10:45 AM - 11:10 AM group, the percent of appropriate crossings increased to 60% and remained high. When the same procedures were implemented in the 11:45 AM - 12:10 PM period, a similar rise in the percent of appropriate crossings occurred. Figure 3 also shows that when the modification procedures produced an increase in the rate of appropriate crossings in the 9:45 AM - 10:10 AM period, the relative rates for the other two time periods with baseline procedures still in effect remained low. When the same procedures were instituted in the 10:45 AM - 11:10 AM period, there was an immediate increase of appropriate crossings (65%), while the rate for the 11:45 AM - 12:10 PM period, with baseline procedures still in effect, remained low. Likewise, instituting modification procedures for the latter time period produced an immediate increase of 61% in the rate of appropriate crossings. During the remainder of the reinforcement procedure, the rates remained relatively high and stable in all three time periods. The percent of inter-observer agreement during the six reliability checks was never below 95%. Reliability was figured on the basis of the percent of inter-observer agreement on the number of appropriate crossings on each corner.

The average number of students who left between 9:45 AM and 10:10 AM was 35. Between 10:45 AM and 11:10 AM the average was 32.4, and between 11:45 AM and 12:10 PM it was 51.7.

DISCUSSION

The results of the experimental procedures support the contention that appropriate street crossing behaviors can be defined, measured and modified. Because of the limited amount of time that was available for this project, some variables were not investigated. It would have been of interest, for example, to demonstrate the differential effects of reinforcement vs. instructions on street crossing behaviors. Past research (Ayllon and Azrin, 1964) indicates that instructions given without providing consequences for the behavior have short term effects.

Another problem is that the time limitations did not allow for the demonstration of the durability of the newly acquired safe street crossing behaviors. The procedures that would have been used here, and the one that is recommended, is to have employed some form of an intermittent schedule of reinforcement. This would be done by gradually dropping reinforcement procedures on a few days, but still providing them on the others. There is a large amount of research which shows that once a response had been conditioned, it can be maintained for long periods by providing intermittent reinforcement (Ferster and Skinner, 1957; Mech, 1953; Fattu, Mech and Auble, 1955; and Long, Hammack, May and Campbell, 1958). Nevertheless, this study demonstrated that efficient and easily applied behavior modification techniques can be used to modify pedestrian behaviors in children.

REFERENCES

Ayllon, T. and Azrin, N. H. Reinforcement and instructions with mental patients. *Journal of Experimental Analysis of Behavior*, 1964, 7, 327-331.

Baer, D. M., Wolf, M. M. and Risley, T. R. Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1968, 1, 91-97.

Fattu, N. A., Mech, E. V. and Auble, D. Partial reinforcement related to free responding in extinction with preschool children. *Journal of Experimental Education*, 1955, 23, 365-368.

Ferster, C. B. and Skinner, B. F. *Schedules of reinforcement*. New York: Appleton-Century-Croft, 1957.

Long, E. R., Hammack, J. T., May, F. and Campbell, B. J. Intermittent reinforcement of operant behavior in children. *Journal of Experimental Analysis of Behavior*, 1958, 1, 315-339.

Mech, E. V. "Resistance to extinction" of the two patterns of verbal reinforcement. *Journal of Experimental Education*, 1953, 22, 155-163.

A PROGRAM OF DELAYED CONSEQUENCES FOR THE MANAGEMENT OF CLASS ATTENDANCE AND DISRUPTIVE CLASSROOM BEHAVIOR OF 124 SPECIAL EDUCATION CHILDREN[1]

hewitt b. clark

A great deal of research has focused upon the elimination of classroom behavior considered to be competing with participation in study activities. The procedures researched have, for the most part, involved consequences which were administered within the classroom setting (e.g., Barrish, Saunders and Wolf, 1969; Birnbrauer, Wolf, Kidder and Tague, 1965; Hall, Fox, Willard, Goldsmith, Emerson, Owen, Davis and Porcia, 1971; Herman and Tramontana, 1971; O'Leary and Becker, 1967; O'Leary, Becker, Evans and Saudargas, 1969; Schmidt and Ulrich, 1969; Thomas, Becker and Armstrong, 1968; and Wolf, Giles and Hall, 1968). The management of classroom behaviors through the use of delayed consequences administered in home settings has been researched by Bailey, Wolf and Phillips (1970) and McKenzie, Clark, Wolf, Kothera and Benson (1968). The present study was designed to extend the results of the McKenzie et al. study and the Bailey et al. study by examining the use of delayed consequences in the management of class attendance and classroom disruptive behavior of a relatively large population of children. The children, numbering 124, attended a special education school and lived in cottages at a residential facility for mildly and moderately retarded children. The delayed consequences for school attendance and disruptive classroom behavior were differentially applied each evening by the cottage-parents. The present study involved two experiments. The first was designed to evaluate the program of delayed consequences on class attendance.

---

[1] The author wishes to express his gratitude to Miss Donna Mae Ida, who had provided continual assistance of a superior quality in this and many other research projects. The author would also like to give a special thanks to Mrs. Sandra Boyd and Miss Susan Brothers for their help and cooperation in conducting this research in their classrooms. This project would not have been possible without the cooperation and assistance of the personnel of the Education Unit at Western Carolina Center. To them, I am indebted, as well as to the Office of Mental Health which provided the support for this project.

GENERAL METHOD

The subject population of this research consisted of 124 children, ages 9-16, who had been institutionalized due to such circumstances as:

(1) their having a history of being too disruptive for public school special education classes;

(2) their community public school not having a special education program; or

(3) the state courts having no alternative facilities available to receive children of these ages who have come under court jurisdiction.

At the institution, the children resided in four cottages. Each child was scheduled to attend six hours of special education classes each day, Monday through Friday. The classes were held in a school building separate from the cottages. There were 14 special education teachers in the school, each with a different curriculum area (e.g., reading, numbers, communication skills, community living, home economics and prevocational training). The children were to go to a different curriculum area class each hour. They were allowed to leave class on the hour and were required to be in their next class by ten minutes after the hour in order to be counted as "present."

In both the school and the cottages, the children were on a *token system*. Within the school the children received plastic strips and praise immediately contingent upon appropriate classroom and academic behavior. At the end of each class period the number of tokens earned was marked in a *token book* carried by each child. The marks in the token books were transferred into each child's *token bank account* daily. A child could take out of his bank account whatever quantity of his token coins he wanted to spend. The token coins could be exchanged for a variety of evening activities, such as dances, movies and organized recreation. These coins could also be exchanged for spending money or used to purchase clothing and a variety of items from a novelty store.

Each evening there was a meeting of all residents in each cottage. Each child who attended this meeting received 25 tokens. Each child who missed the meeting was fined 100 tokens and was placed in bed for the rest of the night when he was located. The evening meeting was used to inform the children of what activities were scheduled for the evening.

EXPERIMENT I

This experiment was designed to evaluate the use of the program of delayed consequences for the management of class attendance. A sample, although not an experimental analysis, was also provided in this experiment of the program's effectiveness in the management of disruptive classroom behavior.

METHOD

Measurement of Class Attendance

A record of class attendance was made each day by each teacher on an *Attendance Form*. This form contained a listing of all the children's names in alphabetical order and two columns in which marks could be made beside each name. For each class, the teachers would check roll at ten minutes after the hour and indicate each child's presence by placing a check in the first column next to the child's name. Only those children who arrived in class before ten minutes after the hour and remained throughout the hour were counted as "present."

Once each day, the reliability of each teacher's attendance record (Attendance Form) was checked. This was accomplished by having an observer drop by each teacher's classroom at a different class hour each day and make an independent record of the children "present." The teacher's record and the observer's record were compared to determine the number of agreements (i.e., both the teacher and the observer had checked the child as present) and the number of disagreements (i.e., either the observer had not checked a child and the teacher had or vice versa). A reliability estimate was calculated across all teachers by dividing the total number of agreements by the total number of agreements plus disagreements. The mean reliability for the teachers' attendance records was 98%.

After the last class of the school day all the Attendance Forms were collected and the attendance information collated for each child. The collation was used to obtain a percentage figure for each child indicating the proportion of his classes he attended. The percentage figures were corrected to compensate for excused absences.

Measurement of Disruptive Classroom Behavior

To provide an estimate of the effect of the treatment program on the rate of disruptive classroom behavior, measurements were made in one classroom (11:00 math class) of *out-of-seat behavior*. Due to the lack of observer staff, these observations were done only under the first two conditions of the study, and then only on a few days of both conditions. On those days on which a classroom observation was taken, two observers made a 15-minute sample of the number of 10-second intervals in which one or more children were out of their seat without permission. The observer's records were later compared at each interval where one or the other observer had scored an out-of-seat behavior. The occurrence reliability was 82% (i.e., number of intervals of agreement that the behavior occurred divided by the number of intervals of agreement and disagreement).

PROCEDURES AND RESULTS

Figure 1 shows the average percent of classes attended by all 124 students and a sample of the rate of disruptive behavior displayed within one class. The left ordinate provides the percentage scale for class attendance. Each dot connected by the solid line shows the average percentage of classes attended during a school day by the 124 students.

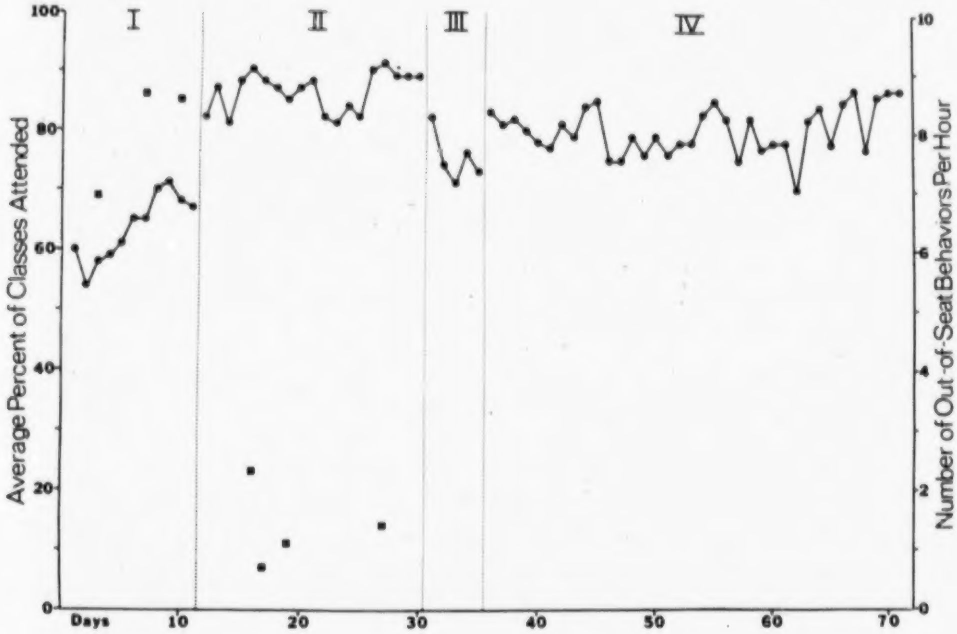


Fig. I. Average percent of classes attended each day are shown as dots connected by the solid lines for each of the four conditions of Experiment I. The percentage scale for attendance is shown along the left ordinate. The rate per hour of out-of-seat behavior for one class is shown for seven days of observation taken during the first two conditions of Experiment I. These rates are represented by the squares and the rate scale is along the right ordinate.

The percentage was calculated by dividing the number of classes attended (including excused absences) by the total number classes available (6 times 124), and multiplying the result by 100 to provide a percentage figure. The right ordinate of Figure 1 provides a scale for the rate per hour of out-of-seat behavior. The squares represent the rate of out-of-seat behaviors observed in the one class on seven different days of observation taken during the first two conditions of the study.

Under the first condition there were no specific contingencies in effect for school attendance or disruptive behavior. An average of 63% of the classes were attended under this baseline condition. The rate of out-of-seat behavior averaged approximately 8 per hour.

At the evening meeting prior to the second condition, the children of each cottage were told about a new program. They were told about a set of new *school rules* (Table 1). These rules were discussed with the children and they were informed that the following set of consequences would be in effect as of the next day:

a▶ If you go to all your classes and do not violate any school rules, (i.e., do not emit disruptive behavior) you will receive 100 bonus tokens at the evening meeting.

b▶ You will be fined 25 tokens for each class you miss. *(unexcused absence)*

c▶ You will be fined 10 tokens for each school rule you violate during the day.

d▶ If you miss many of your classes, you will be placed in bed for the entire night. *(bed restriction)*

e▶ If you violate a large number of rules you will be placed in bed for the entire night. *(bed restriction)*

The program was outlined and discussed with the teachers the next morning prior to classes on the first day of this condition. The specific procedures of fining for rule violations were described and modeled. The fining for rule violations was done in the following manner. Each time the teacher saw a child violate a rule, she was to say the child's name, indicate the fine, and give a brief verbal description of the behavior fined, (e.g., "Mary, ten token fine for talking without permission;" "John, ten token fine for not working.") Immediately following this, the teacher was to put a mark ("*1*") on the Attendance Form by the child's name. Each of these marks meant a ten token fine.

TABLE 1

A LIST OF THE SCHOOL RULES

1. WHEN CLASS BEGINS, YOU ARE TO BE SEATED IN YOUR CHAIR AND QUIET.
2. YOU ARE TO REMAIN IN YOUR CHAIR UNLESS THE TEACHER GIVES YOU PERMISSION TO GET UP.
3. IF YOU WANT TO SAY SOMETHING, RAISE YOUR HAND AND WAIT UNTIL THE TEACHER CALLS ON YOU.
4. YOU ARE NOT TO MAKE NOISE OR DO THINGS WHICH MIGHT BOTHER YOUR CLASSMATES.
5. YOU ARE TO PARTICIPATE IN *ALL* CLASS ACTIVITIES.
6. YOU ARE NOT TO TEASE OR FIGHT WITH OTHER CHILDREN.
7. THERE IS TO BE NO FIGHTING, YELLING, OR RUNNING IN THE HALLS OF THE SCHOOL.
8. YOU ARE NOT TO DISTURB OTHER CLASSES.
9. YOU ARE TO GO TO ALL YOUR CLASSES.
10. YOU ARE NOT TO BE IN THE CANTEEN EXCEPT BETWEEN CLASSES.
11. SMOKING, EATING, DRINKING, OR LISTENING TO RADIOS ARE NOT ALLOWED IN THE SCHOOL BUILDING. THESE ACTIVITIES ARE PERMITTED IN THE CANTEEN.
12. YOU ARE NOT TO SAY CURSE WORDS OR TALK BACK TO TEACHERS.
13. YOU ARE NOT TO DESTROY ANY SCHOOL PROPERTY.
14. ALL SUPPLIES ARE TO BE PUT IN THEIR PROPER PLACE BEFORE LEAVING THE ROOM.

At each evening meeting during the second condition the following consequences were administered:

- ●a. The children who had gone to all of their classes and had not received any marks for rule violations were called to the front of the room and were given a bonus of 100 tokens.
- ●b. The names and amounts of fines were read of those children who had received fines due to either class absences or rule violations. (Prior to the meeting, these fines had been subtracted from the children's bank accounts.)
- ●c. In each cottage the two children with the lowest percentage of class attendance were restricted to their beds for the rest of the night.
- ●d. In each cottage the two children with the largest number of fines for rule violations were restricted to their beds for the rest of the night.

With the delayed consequences for attendance and rule violations in effect, the percentage of class attendance increased to an average of 87% from a baseline average of 63%. The sample of rule violations, represented by the squares, showed a decrease in the rate of out-of-seat behavior from an average of eight per hour under baseline to an average of 1.4 per hour on the program of delayed consequences. This decrease occurred despite an increase in the average daily attendance to this class. That is, an average of 5.7 students attended this class under the baseline and an average of 9.2 attended under the second condition.

After 19 days under this program of delayed consequences for both school attendance and rule violations, the conditions were changed such that the delayed consequences for school attendance were discontinued while the consequences for rule violations were maintained. At the evening meeting prior to this next condition, the children were told that they should continue to go to all their classes, but that they would not receive any bonuses, fines, or bed restrictions for class attendance. They were also informed that consequences would still be in effect for rule violations. Thus, at the evening meeting during the third condition the following consequences were in effect:

- ●a. The children who had not received any marks for rule violations were called to the front of the room and given a bonus of 50 tokens.
- ●b. The names and the amounts of fines were read of those children who received fines because of rule violations.

c. In each cottage the two children with the largest number of rule violations were restricted to their beds for the rest of the night.

When the specific contingencies for class attendance were discontinued under this third condition for a five day period, the percentage of classes attended decreased to an average of 75%. Class attendance was again increased when the contingencies for class attendance were reinstated under the fourth condition. Although class attendance did increase to 81% under this fourth condition, it did not average as high a percentage as maintained under the second condition. However, the fourth condition did maintain a higher average than during the first or third conditions in which no specific consequences were scheduled.

EXPERIMENT II

This second experiment was conducted to provide an analysis of the effectiveness of the program in the management of disruptive classroom behavior.

METHOD

Three weeks after the completion of Experiment I, Experiment II was begun. The program outlined in Experiment I remained in effect throughout Experiment II.

A multiple baseline design across the 11 o'clock and the 1 o'clock art classes provided the basis for the analysis in Experiment II. The same teacher had both art classes. The 11 o'clock class had an average attendance of 10 students per day and the 1 o'clock had an average of 12 students per day.

Measurement of Classroom Behavior

A 10-second interval system was employed to measure the percentage of time in which one or more children violated a rule by either talking without permission or getting out of their seat without permission. The teacher's rate of social praise, tokens, and fining was also measured. The observers took an observation each day for 30 minutes in each class. The teacher's rate of praise and tokens was observed for one minute, the next minute the teacher's rate of fining was observed, and the following minute involved the measures of the children's out-of-seat and talking-out behavior. Thus, over the 30 minutes, a 10-minute sample of the children's violation of out-of-seat behavior and talking-out behavior was obtained. The average occurrence reliability for each measure was as follows: teacher reinforcement=93%; teacher fining=98%; talking-out=89%; and out-of-seat=91%.

PROCEDURES AND RESULTS

Figure 2 presents the percentage of time (i.e., 10-second intervals) in which one or more children were talking without permission. This measure is shown in the upper graph for the 1 o'clock (PM) class and in the lower graph for the 11 o'clock (AM) class. The abscissa represents the days of the study.

During each of the first three days of this study, the teacher spent approximately 5 to 10 minutes at the start of both class periods discussing the school rules with the children. The rules were posted on

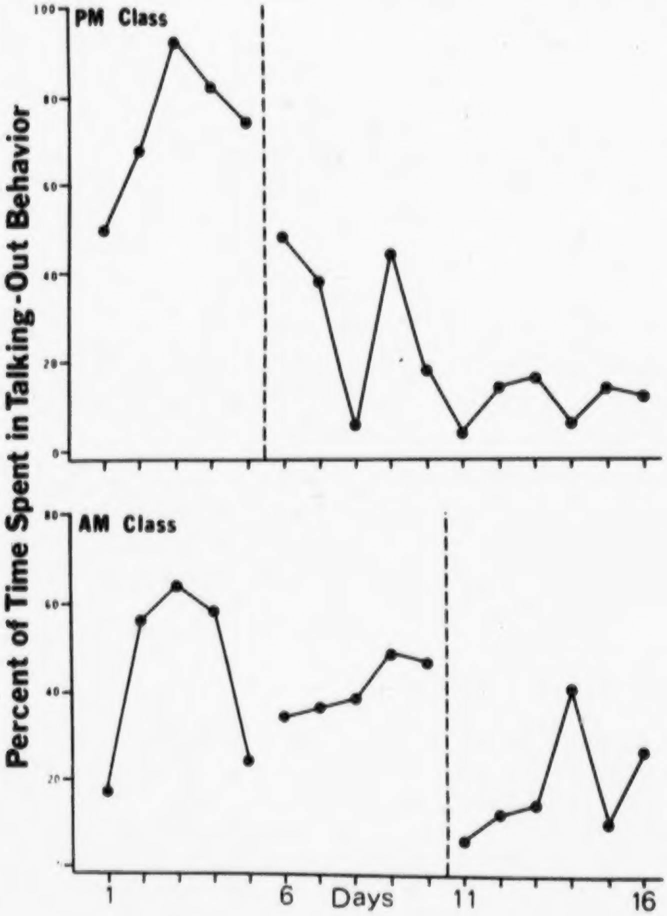


Fig. 2. Percent of 10-second intervals in which one or more students were talking-out without permission is shown for two classes (PM class or 1 o'clock art class and AM class or 11 o'clock art class).

a big chart in the classroom such that they could be seen by the children. These school rules had been in effect for nearly three months at the start of this second experiment.

Under the initial condition or baseline of this study (days 1-5 of the PM class and days 1-10 of the AM class), the teacher attempted to reinforce with social praise and tokens appropriate classroom and academic behaviors, and ignored rule violations and other inappropriate behaviors. No fines were used under this baseline condition in either class. After five days, the teacher began fining all instances of rule violations in the PM class. As can be seen in Figure 2, the percentage of time in which one or more children were engaging in talking-out without permission decreased to an average of 20% from an average of 74% under the baseline condition. After another five days, the teacher began fining all rule violations occurring in the AM class. This resulted in a decrease in the percentage of time in which the children were talking without permission. Under baseline talking-out occurred 44% of the time and was decreased to 23% of the time under the treatment condition.

The percentage of time in which one or more children were out of their seats without permission was also measured. Figure 3 shows that the results obtained on rule violations involving out-of-seat were similar to those obtained for talking-out violations; however, they were not as clear.

Reductions in the percent of time that children were involved in rule violations resulted despite the fact that teacher reinforcement remained approximately constant under the baseline and treatment conditions for each class. In the PM class, 42% of the teacher intervals observed under the baseline condition included the delivery of reinforcement. During the treatment condition for this class, the teacher delivered reinforcement during 44% of the intervals. In the AM class, the percentages were 38% and 39% for the baseline and treatment conditions, respectively.

DISCUSSION

The results of the present study suggest that delayed consequences can be effective in increasing class attendance and in the reduction of disruptive-classroom behavior. The program of delayed consequences consisted of reinforcement and punishment procedures which were differentially applied each evening based upon the children's school attendance and classroom behavior. These results appear consistent with previous research concerning the management of classroom behavior through the use of *home-based* reinforcement (Bailey et al., 1970; and McKenzie et al., 1968).

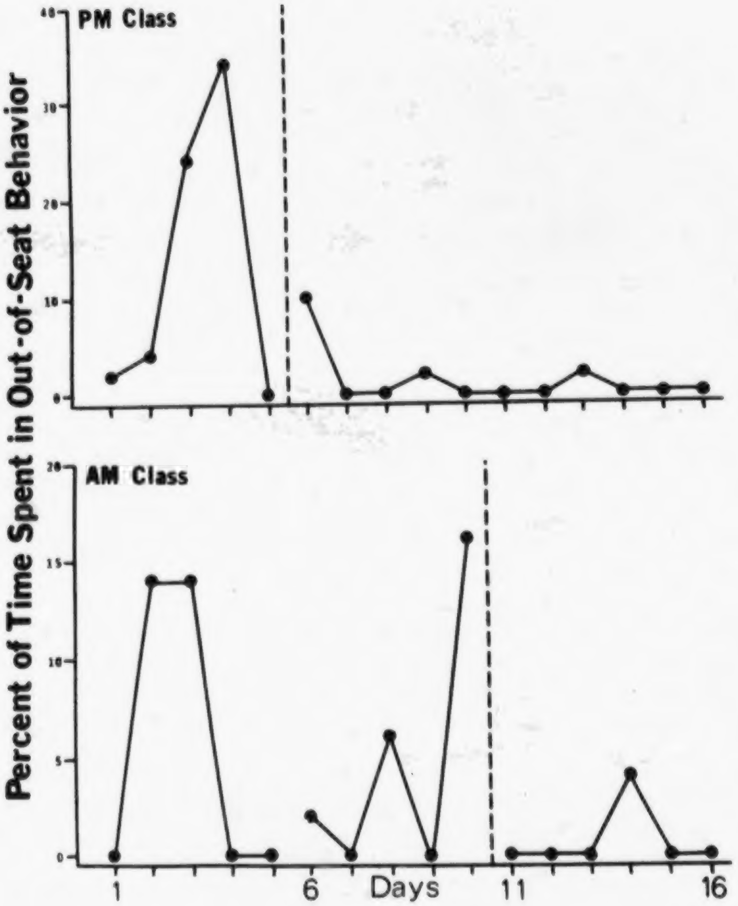


Fig. 3. Percent of 10-second intervals in which one or more students were out-of-seat without permission is shown for two classes. (PM class or 1 o'clock art class and AM class or 11 o'clock art class.)

Although the present study involved the delayed consequences being administered in the children's home cottage, it is likely that the program could be adapted to a normal public school. In a public school setting the consequences could include *reinforcing* activities and materials which are a natural part of the setting (e.g., recess; outdoor play materials; access to art, gym and music classes; access to basketball practice and the gym over the noon hour; or an opportunity to leave school early). Because most elementary schools use self-contained classrooms, the focus of the management program may not be on class attendance, but more likely on classroom academic and non-academic behavior. For example, the teachers of an elementary school might each carry a *Reinforcement and Fine* sheet with them at all times. The teachers would place a mark in the *reinforcement* column next to the student's name each time the teacher socially praised a student for correct answers. Similarly, marking in the *fine* column for each school rule violation. The forms would be collected at a specific time during the day and the marks tallied. Those students with the most reinforcement marks would have access to special activities versus staying in a study hall, for example.

A few possible advantages of such a program over a more traditional *token system* are:

(1) any student can be reinforced and/or fined at any time by any teacher (e.g., between classes, in different classrooms, on field trips);

(2) fining a student does not require that the teacher physically interact with the student to take away *tokens*;

(3) the calculations of the marks earned and lost from the reinforcement and fine form may prove to be more economical in time than counting individual student's tokens; and

(4) the reinforcement and fine forms may provide helpful feedback to teachers (and principals) to assist them in establishing and maintaining a high rate of social praise to their students.

Determining the effectiveness and feasibility of such a program of delayed consequences for regular elementary schools will, of course, require much additional research attention.

REFERENCES

Bailey, J. S., Wolf, M. M. and Phillips, E. L. Home-based reinforcement and the modification of pre-delinquents' classroom behavior. *Journal of Applied Behavior Analysis*, 1970, 3, 223-233.

Barrish, H. H., Saunders, M. and Wolf, M. M. Good behavior game: effects of individual contingencies for group consequence on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis*, 1969, 2, 119-124.

Birnbrauer, J. S., Wolf, M. M., Kidder, J. D. and Tague, C. E. Classroom behavior in retarded pupils with token reinforcement. *Journal of Experimental Child Psychology*, 1965, 2, 219-235.

Hall, R. V., Fox, R., Willard, D., Goldsmith, L., Emerson, M., Owen, M., Davis, F. and Porcia, E. The teacher as observer and experimenter in the modification of disputing and talking-out behaviors. *Journal of Applied Behavior Analysis*, 1971, 4, 141-149.

Herman, S. H. and Tramontana, J. Instructions and group versus indivual reinforcement in modifying disruptive group behavior. *Journal of Applied Behavior Analysis*, 1971, 4, 113-119.

McKenzie, H. S., Clark, M., Wolf, M. M., Kothera, R. and Benson, C. Behavior modification of children with learning disabilities using grades as tokens and allowances as back up reinforcers. *Exceptional Children*, 1968, Summer, 745-752.

O'Leary, K. D. and Becker, W. C. Behavior modification of an adjustment class: a token reinforcement program. *Exceptional Children*, 1967, May, 637-642.

O'Leary, K. D., Becker, W. C., Evans, M. B. and Saudargas, R. A. A token reinforcement program in a public school: a replication and systematic analysis. *Journal of Applied Behavior Analysis*, 1969, 2, 3-13.

Schmidt, F. W. and Ulrich, R. E. Effects of group contingent events upon classroom noise. *Journal of Applied Behavior Analysis*, 1969, 2, 171-179.

Thomas, D. R., Becker, W. C. and Armstrong, M. Production and elimination of disruptive classroom behavior by systematically varying teacher's behavior. *Journal of Applied Behavior Analysis*, 1968, 1, 35-45.

Wolf, M. M., Giles, D. and Hall, R. V. Experiments with token reinforcement in a remedial classroom. *Behavior Research and Therapy*, 1968, 6, 51-64.

⁂

3 SIMPLE PUNISHMENT TECHNIQUES FOR THE CONTROL OF CLASSROOM DISRUPTIONS

**thomas sajwaj; penelope culver
carolyn hall; les lehr**

The control of classroom disruptions has been usually obtained by the use of a token reinforcement system, by the use of differential attention, or by the use of some form of a time-out procedure. In a token reinforcement system, some type of token or point is given for desirable behavior, usually of a typography incompatible with the disruptive behavior. The disruptive behavior is either ignored, punished with a fine (i.e., loss of tokens or points), or punished with the presentation of *minus* points which must be *earned off* to avoid some other form of aversive event (Broden, Hall, Dunlap and Clark, 1970). Similarly, the use of differential attention necessitates relatively high levels of attention and praise for desirable behaviors incompatible with the disruptive behaviors. In addition, the disruptive behaviors would be ignored (Madsen, Becker and Thomas, 1968). The use of time-out usually involves moving the child out of the classroom situation to a setting which is ideally less reinforcing to the child whenever he engages in the disruptive behaviors. This latter setting might be a barren room, the principal's office, the hallway, or a chair in the back of the classroom itself (Zeilberger, Sampen and Sloane, 1968).

All of these techniques have been shown to be effective in controlling disruptive behaviors. However, each technique has some encumbrance in its utilization which will limit or eliminate its use in a given situation. Token systems necessitate some mechanics for giving tokens (e.g., points on pads of papers, stars on a card) with some clerical auditing involved. In addition, backup reinforcers need to be developed and time taken to purchase these reinforcers. This effort can involve a major restructuring and reorganization of the teacher's classroom, an effort she may be unwilling to make. The use of differential attention by some teachers will involve a major modification of their typical classroom behavior. Again, some teachers will be unwilling to make this change. Others will not be able to make it. Also, there are some disruptive behaviors which cannot be ignored because of their very nature (e.g., severe aggression). Relative to time-out, convenient and practical areas to send the child may be unavailable. Some principals prohibit the use of their offices or of the hallways, and unused rooms might be nonexistent. Thus, it is quite possible that none of these techniques are practical in a given classroom situation. At this point, the teacher must have accessible other techniques which are more suitable and practical for her situation.

Hall, Axelrod, Foundopoulos, Shellman, Campbell and Cranston (1971) have provided four such simple techniques. These involved the teacher pointing her finger and saying *no*, the removal of a colored slip bearing the child's name, remaining after school, and after-school tutoring. All were effective in controlling undesirable classroom behavior. More importantly, they involved less encumbrances to their usage in class than token systems, differential attention and time-out.

The present paper presents three other types of punishment procedures which were effective in reducing disturbances and which involved minimal encumbrances and effort on the teacher's part.

EXPERIMENT I[1]

Children

This study was conducted in a regular third grade math class comprised of twenty-four "slow learners." Class met daily for one hour. The class, as a whole, tended to be disruptive with talk-out and out-of-seat behaviors being most frequent. Although the program was implemented with the whole class, data was obtained only on two boys, Albert and Harry, whom the teacher regarded as the most disruptive in the class.

Measurement and Reliability

The teacher recorded on a slip of paper each talk-out and out-of-seat behavior that occurred for Albert and Harry during the last twenty minutes of the class period. The last twenty minutes was selected because it tended to be their worst behaved period. Talk-outs were defined as any audible verbalization without permission. Out-of-seat behavior was defined as being away from the child's assigned desk without permission. Reliability checks were taken in each condition by having a volunteer take independent counts. Reliability estimates were computed by dividing the smaller by the larger count. The range of these estimates was from 81% to 100%, with a mean of 92%.

Program

After eight days of baseline observation under the normal conditions of the class, the teacher initiated a *five star general* program for the entire class. Each child's name was posted on a large chart in the front of the class. If he did not talk-out or was not out-of-seat beyond a minimum number per class period, he received a star by his name on the chart. This minimum number was varied from one to three. When a child accumulated five such gold stars, he was awarded a *five star general* badge to wear for one day. This badge was a simple small bow of red, white, and blue ribbon with a golden metal button, which was worn on the shirt or blouse. When a child wore this badge, the other children were told to *honor* the wearer by saluting him whenever they passed him in school. This salute consisted of a raised right hand with the thumb and middle finger touching and forming a circle. When a child had earned a badge on two occasions, he was awarded a special letter of commendation by the teacher which he was to show to his homeroom teacher, the principal, and his parents. This commendation stated in flowery script and in formal language that the child had been excelling in good citizenship in his math class. After eight days of this program, baseline conditions were recovered for an additional eight days. The program was then reinstated.

---

[1] Reprints may be obtained from the first author at the Department of Psychiatry, University of Mississippi Medical Center, Jackson, Mississippi 39216. Experiments I and II were partially supported by The Avery Fund for Research in The Behavioral Sciences.

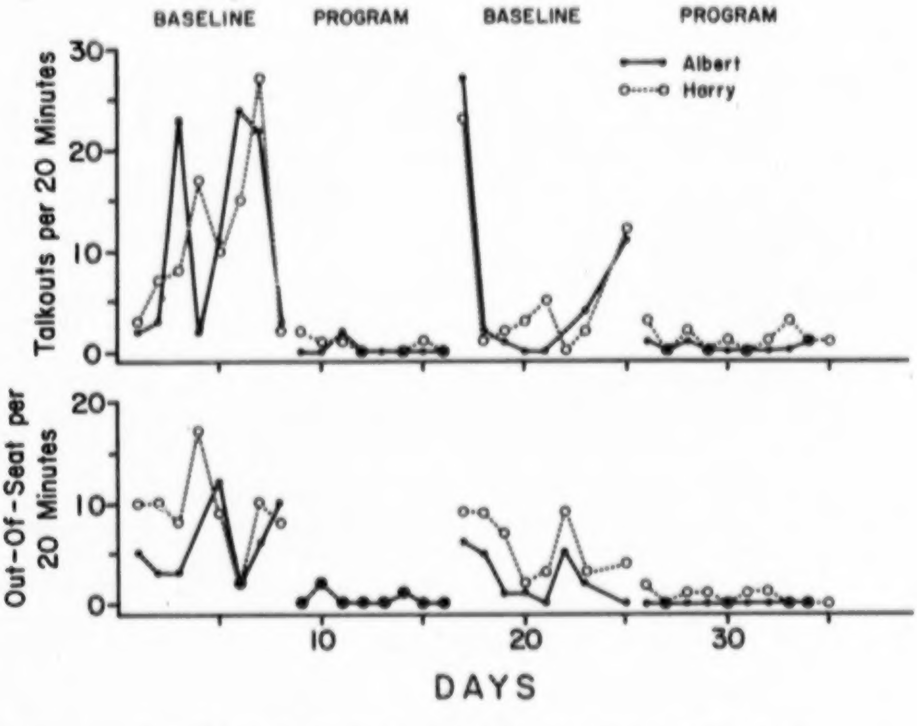


Fig. 1. Number of talk-outs and out-of-seat episodes in twenty minutes for Albert and Harry in Experiment 1 as a function of baseline conditions (no systematic program) and of golden star and general badges for minimal disruptive behaviors (program).

Results

Figure 1 shows the rate of talk-out and out-of-seat behaviors per twenty minutes for Albert and Harry. During the initial baseline condition, Albert and Harry each talked about ten to fifteen times per twenty minutes and were out-of-seat between five and ten times. During the initial use of the *five star general* program, talk-out and out-of-seat behavior dropped to virtually zero. The return to baseline conditions prompted an increase in both behaviors in both boys. The reinstatement of the program recovered the very low rates of these disruptive behaviors. Subjectively, the teacher reported that the program was as effective with the other children in the class. The class atmosphere improved, and the children became more productive.

EXPERIMENT II

Children

A fifth grade class comprised of twenty-four children of average intelligence attended a one hour class. Data on poor sitting posture (slouching) were collected on the entire class. In addition, talk-out data were obtained on one boy, Charles, because the teacher felt that he had the highest level of disruptive behavior in the class.

Measurement and Reliability

For slouching, the teacher made four checks during the class at ten minute intervals. The number of children not sitting correctly was recorded each time. At the end of the class, an average of these four checks was computed. Sitting erect with feet flat on the floor in front of the desk and with both arms on the desk top was considered proper sitting posture. Any deviation from this was considered slouching behavior. Students who were out-of-their-seats for any reason or who were leaning down to get books out of their desks were not counted as slouching. Charles' talk-outs were tallied during a 20-minute period by the teacher marking on a slip of paper. A talk-out was defined as any audible talking sound made without permission. Reliability checks were made on both categories of behavior in each condition of the study by having an independent observer record these behaviors.

Reliability estimates were computed by dividing the smaller score by the larger score. For slouching behavior, these estimates ranged from 87% to 100%, with a mean of 92%. For talk-outs, the range was from 86% to 100% with a mean of 96%.

Program

BASELINE observations were made for five days under typical classroom conditions. LOSS I conditions were then implemented. The children were paired according to rough subjective estimates of their disruptive behavior, with the two best together, two worst together, etc. Each child was given 100 points at the start of each class period. Each time a child talked-out or slouched, one point was subtracted from his total. Running totals were conspicuously displayed at the front of the class. Toward the end of the class period, the child in each pair with the most points was allowed ten minutes of free time to do as he pleased. The loser in each pair was given more classwork for that ten minutes. The LOSS II condition was identical with the LOSS I condition except that free time was not awarded to winners. Instead, winners in each pair were only announced at the end of each class period and attention drawn to their accomplishment. This was done to allow a comparison of the effects of public display and competition with the effects of the free time privilege.

LOSS I, LOSS II, and BASELINE conditions were then alternated to isolate the effect of each type of condition.

Results

Figure 2 gives the daily rates of slouching behavior for the whole class and of talk-outs per twenty minutes for Charles. During BASELINE conditions, rates of both behaviors were relatively high. About ten or more children (40% plus) would be slouching at any one time. Charles' talk-outs were above ten for the 20-minute period. During LOSS I conditions (public display and competition, with free time privilege), rates of both behaviors were virtually zero. However, during LOSS II conditions (public display and competition alone) slouching rate would initially be low but would then slowly increase. This slow increase occurred on all three exposures of the class to the public display and competition factors alone. Charles' talk-out rate during LOSS II conditions was roughly intermediate between BASELINE and LOSS I conditions.

It is clear that public display and competition did reduce slouching and talk-outs, at least, temporarily. However, maximal reductions were obtained by the inclusion of the free time reinforcer with public display and competition. It would appear that the free time privilege is the more potent factor. This cannot be asserted unequivocally since the free time privilege was not presented alone to determine its effectiveness independent of public display and competition.

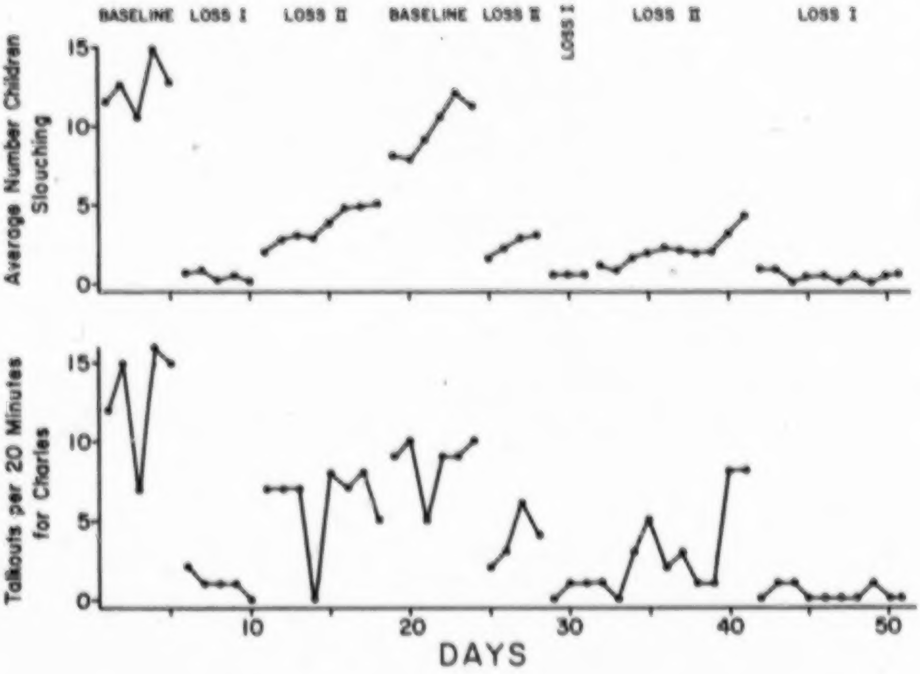


Fig. 2. Average number of children slouching at any time and Charles' talk-outs per twenty minutes in Experiment II as a function of untreated (BASELINE) conditions, public display and competition with free time privileges (LOSS I), and public display and competition alone (LOSS II).

EXPERIMENT III[2]

Child

Mark was a six year old, retarded child with a history of convulsions, vision difficulties, tantrums, and enuresis. The convulsions were controlled by medication. His speech, social behavior, and play behavior were poorly developed for his age. He could not participate in the regular or special education programs of the public schools. Instead, he was placed in an experimental preschool for retarded and/or emotionally disturbed children.

During the course of his participation in this program, a serious throwing problem developed. Mark would throw large objects, usually toys, across the classroom. This throwing was rarely aimed at anyone, but there was a real danger to the other children and to the teacher.

Measurement and Reliability

An observer recorded the frequency of Mark's throwing by checking off each ten second interval in which any throwing occurred (Allen, Hart, Buell, Harris and Wolf, 1964). Throwing was defined as picking up any object with his hands and throwing it so that it landed more than one foot away from him. Mark's throwing was recorded for twenty minutes during free play each morning.

Reliability checks were made by having an independent observer record throwing behavior with the regular observer on fourteen occasions during the first two conditions of the study. Reliability estimates were obtained by comparing each interval and counting an agreement if both observers scored throwing. If only one observer scored throwing, a disagreement was scored. If neither scored throwing, neither an agreement nor disagreement was scored. The number of agreements was then divided by the sum of the agreements plus disagreements. These ranged from 57% to 100%, with a mean of 92%. The lower estimates were obtained in the reprimand condition. However, all observers in this condition agreed that overall rate of throwing was reduced sharply.

---

[2] Experiment III was supported by the United States Office of Education Grant #0762(032)46-32-66-358 to Dr. S. W. Bijou. The authors wish to express their appreciation to Sandra Twardosz for her generous cooperation in Experiment III.

Program

During BASELINE conditions, throwing was ignored. However, the teacher made efforts to keep the other children out of the range and direction of his throwing. During REPRIMAND conditions the teacher would immediately come over to Mark after each episode of throwing. She would grab his shoulders, look him straight in the face, and sternly say, "Mark, I don't want you to do that!" She then released him and returned to her duties. This REPRIMAND procedure took about three to four seconds. After the first few reprimands, Mark began shaking his shoulders in an apparent effort to make the teacher release her grip on his shoulders. The teacher was then instructed to keep her grip until he clearly stopped shaking. She had to hold him for as long as 15 seconds, but the shaking disappeared rapidly. After fourteen days of this REPRIMAND procedure, BASELINE conditions were restored where throwing was ignored. After this condition, the REPRIMAND procedure was reinstated for three days. Mark was then terminated from the preschool because of very poor toilet training.

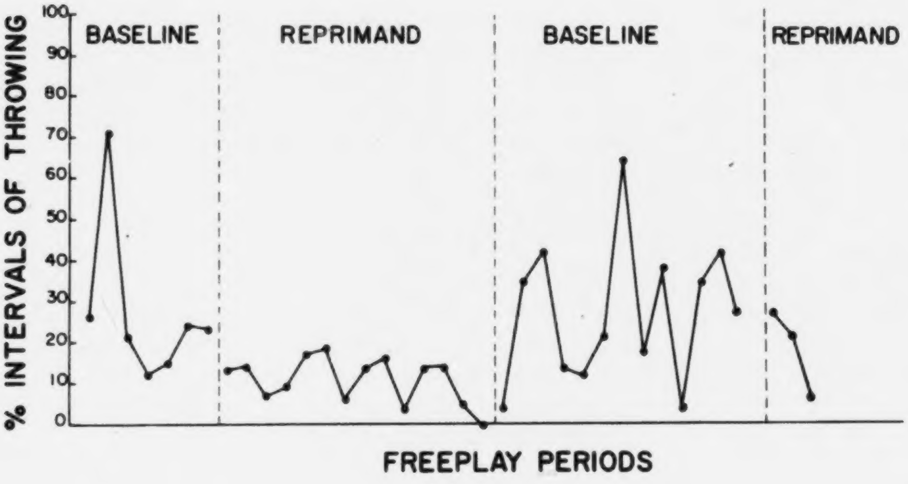


Fig. 3. Percentage of throwing intervals in Experiment III as a function of ignoring (BASELINE) and REPRIMAND.

Results

During BASELINE conditions, throwing occurred in about 20% to 30% of the free play period. During the initial REPRIMAND condition, throwing rate slowly fell to zero. More strikingly, the throwing of large objects (e.g., large wooden blocks) disappeared very rapidly during the first three to five days of the REPRIMAND condition. In addition, the distance of throws fell rapidly. Consequently, during the latter part of the first REPRIMAND condition he was throwing only small objects (e.g., tinker toy sticks) for only about one to two feet, whereas during the BASELINE conditions he threw large objects across the classroom. This effectiveness was recovered during the abbreviated second REPRIMAND condition.

CONCLUSIONS AND DISCUSSION

These experiments demonstrate three simple punishment techniques to be effective in controlling classroom disturbances. In Experiment I, disruptions beyond a given minimal number were punished with the loss of a gold star with the consequent loss or delay in earning the *five star general* badge and its accompanying attention and prestige. In Experiment II, slouching and talk-outs were punished by the loss of points. Losing more points than one's competitive partner resulted in doing ten more minutes of academic work. In Experiment III, the social reprimand with the mild restraint of holding was effective in reducing dangerous throwing. More importantly, each of these three techniques was simple and practical to use. Backup reinforcers and clerical auditing, as needed in a token system, were minimized. A high level of attention for desirable behavior, as needed in a differential attention procedure, was unnecessary. The need for facilities outside the classroom, as often needed with a time-out procedure, was very minimal. Thus, a high level of effectiveness in the control of classroom disruptions was obtained at a relatively low response cost to the teacher. Such an increased efficiency is obviously conducive to the increased usage of punishment techniques and behavior modification procedures, in general, in educational practice.

The procedures in Experiment I are similar to those used by Osborne (1969). He withdrew free time from academic work for any occurrence of out-of-seat behavior in deaf children. Experiment I set a maximal number of disturbances after which the opportunity to obtain a gold star was lost. This, in turn, delayed the opportunity to earn the *five star general* badge and the accompanying system of attention and prestige. The major difference between these two procedures is the nature of the reinforcer lost.

The procedures of Experiment II are similar to several techniques used previously. Most similar is the *Good Behavior Game* (Barrish, Saunders and Wolf, 1969). Their class was divided into two teams. Marks were given for each occurrence of out-of-seat and talk-out behavior. The team with the fewest marks earned several classroom privileges (e.g., free time). In Experiment II points were lost, while they were given by Barrish et al., an apparently trivial procedural difference. Experiment II suggests, more importantly, that the public display and competition were relatively less important than the gain of the classroom privileges. However, Barrish et al. used large groups competing against each other, whereas Experiment II used individuals. It may very well be that competition and peer pressures were more relevant factors in the Barrish et al. form of *The Good Behavior Game*.

The simple, yet effective reprimand of Experiment III is similar to the pointed finger and *no* reported by Hall et al. (1971) which was highly effective in reducing biting and pinching in a deaf girl. Also, O'Leary, Kaufman, Kass and Drabman (1970) found soft reprimands to be effective in reducing classroom disruptions. The prime advantage of a reprimand procedure is its utter simplicity. It is the simplest and least encumbered procedure of all those cited in this paper. However, reprimands and commands do not always function as punishers. Thomas, Becker and Armstrong (1968) have found these to help maintain undesirable behavior.

These three simple experiments, together with those studies cited above, suggest that there is a wide range of very flexible procedures which can be used to control classroom disruptions. One does not have to institute a token economy or to utilize a time-out area to maintain order in a class. Further, efficiency may be improved since these techniques generally have fewer encumbrances than token economies and time-out procedures.

Finally, it should be noted that techniques such as *The Good Behavior Game* are very difficult to classify as a positive reinforcement or punishment procedure. They inherently seem to be both, so that a separation of reinforcing and punishing functions is impossible.

REFERENCES

Allen, D. E., Hart, B. M., Buell, J. S., Harris, F. R. and Wolf, M. M. Effects of social reinforcement on isolate behavior of a nursery school child. *Child Development*, 1964, 35, 511-518.

Barrish, H. H., Saunders, M. and Wolf, M. M. Good behavior game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis*, 1969, 2, 119-124.

Broden, M., Hall, R. V., Dunlap, A. and Clark, R. Effects of teacher attention and a token reinforcement system in a junior high school special education class. *Exceptional Children*, 1970, 36, 341-349.

Hall, R. V., Axelrod, S., Foundopoulos, M., Shellman, J., Campbell, R. A. and Cranston, S. The effective use of punishment to modify behavior in the classroom. *Educational Technology*, 1971, 11, 24-26.

Madsen, C. H., Becker, W. C. and Thomas, D. R. Rules, praise, and ignoring: elements of elementary classroom control. *Journal of Applied Behavior Analysis*, 1969, 1, 139-150.

O'Leary, K. D., Kaufman, K. F., Kass, R. E. and Drabman, R. S. The effects of loud and soft experiments on the behavior of disruptive students. *Exceptional Children*, 1970, 37, 145-155.

Osborne, J. G. Free-time as a reinforcer in the management of classroom behavior. *Journal of Applied Behavior Analysis*, 1969, 2, 113-118.

Thomas, D. R., Becker, W. C. and Armstrong, M. Production and elimination of disruptive classroom behavior by systematically varying teachers' behavior. *Journal of Applied Behavior Analysis*, 1968, 1, 35-45.

Zeilberger, J., Sampen, S. E. and Sloane, H. Modification of a child's problem behaviors in the home with the mother as therapist. *Journal of Applied Behavior Analysis*, 1968, 1, 47-53.

USING REINFORCEMENT CONTINGENCIES & TEACHING AIDS TO ALTER SUBTRACTION PERFORMANCE OF CHILDREN WITH LEARNING DISABILITIES[1]

deborah d. smith
tom c. lovitt
john d. kidder

The procedures used by teachers to change the mathematics behaviors of pupils are either contingently or non-contingently associated with the responses of children. Two projects are described in this report. Both deal with the subtraction performance of learning disabled children. The pupil in PROJECT I was eleven; the student in PROJECT II was ten.

In the first project a contingency was scheduled that positively influenced subtraction performance. During certain phases, a withdraw of reinforcement contingency was imposed--minutes of recess time were taken away contingent on incorrect answers. A multiple baseline design was used.

In the second project a series of non-contingent events altered subtraction performance. During some conditions of the study traditionally used teaching aids--*paper clips*, *abacus*, *cuisinare rods*--were used by the pupil to assist in computing subtraction facts. In PROJECT II a series of ABA designs was used.

When a pupil is not responding correctly or efficiently to academic materials, the teacher must select a remediation tactic. Broadly speaking, two remediation strategies are available--teaching routines and contingency management procedures.

Teaching routines are arranged by teachers on a non-contingent basis. They are scheduled to occur regardless of the quality or rate of the pupil's responding. Instructions, for example, were used effectively

---

[1] This research was partially supported by a grant from the United States Office of Education, Bureau of Education for the Handicapped, Grant No. OEG-0-70-3-916(607).

by Lovitt and Smith (1972) to increase the verbal output of a child. Lovitt and Curtiss (1968) showed that a tactic such as verbalizing mathematics problems before responding increased accuracy. In both studies the teaching routine was non-contingently programmed. Modeling and cuing are other instructional techniques scheduled on a non-contingent basis. Many teaching aids such as *cuisinare rods*, *number lines*, and the *abacus* are scheduled non-contingently.

By contrast, reinforcement procedures are arranged on a contingent basis. The prescribed circumstance will occur only if the pupil behaves in a certain way. Reinforcement procedures have been used to attenuate certain behaviors as exemplified in a study by Becker, Madsen, Arnold and Thomas (1967), when they successfully reduced classroom behavior problems. Reinforcement contingencies have also been used to increase certain behaviors. For example, Staats and Butterfield (1965) demonstrated that certain reading behaviors could be accelerated.

The variations in which contingencies can be arranged are many. In academic areas, for example, where correct and incorrect answers are generally tallied, a contingency could be associated with correct answers, incorrect answers, or both. The ratio of such a contingency has infinite variations. The reward or consequence used could range from praise or candy to a special privilege or free time.

Although more behavioral research has been conducted to study the effects of contingency management procedures than teaching routines, teachers are generally more apt to employ the latter method than the former. Some teachers probably use 25 teaching routines for each contingency they arrange. Most teachers, in their response to a pupil's academic failure, initially select a teaching routine to alter performance; if that event proves ineffective, they select a different one. If the child, for instance, can't solve the problems in one text, a model is provided, cuing is offered, flash cards are used, a number line is provided.

Quixotic searches for perfect teaching procedures could require months before the perfect match is found. Other approaches are required to assist some students; for many pupils could perform the behavior requested by the teacher, but are unwilling. Some pupils are more reinforced by staring out windows than completing their math assignments. Others seem to enjoy playing games with their teachers. Still other pupils see no point in doing their math. With children who actively resist instruction, teachers could alter non-contingent routines until dooms day to no avail.

Conversely, is the rare teacher who has an affinity for utilizing contingencies to aid in effecting the academic proficiencies of her pupils. This teacher believes that anything can be taught if only the effective reinforcer can be found and contingently arranged with the desired behavior. As her counterpart impetuously searches for the ideal teaching routine, so the contingency prone teacher chases the elusive motivator. Such a teacher, attempting to instruct a child to subtract might initially arrange a setting that provides the pupil with one minute of play time for each ten correct answers. If that failed, the ratio might be reduced. If those changes also failed to stimulate correct responding, the consequence itself might be altered from music to television, to praise, to candy. It could be, however, that the child has no part of the required behavior in his repertory. In such a case, the manager could arrange countless contingencies without affecting the pupil's behavior.

Certainly, in the first instance where the child could perform the desired behavior but wouldn't, some reinforcement contingency should have been arranged, rather than switching from one teaching procedure to another. Likewise, in the second example where the child could not perform the desired behavior and didn't, a teaching routine should have been set up rather than trying to alter behavior with contingency management.

Following are two projects; both are concerned with increasing subtraction proficiency in children. The first project demonstrates how, under certain conditions, contingency management procedures can effectively influence a pupil's responding. By contrast, the second project illustrates how, under different circumstances, teaching routines, in the form of instructional aids, can positively alter accuracy.

*PROJECT I

METHOD

Subject and Setting

The student in this study was an eleven-year-old girl, Susie, referred to the Experimental Education Unit (EEU), University of Washington, because of reading difficulties. For the purposes of another experiment, Susie was given two achievement tests at the beginning of the school year, the Wide Range Achievement Test (WRAT) (Jastak, Bijou and Jastak, 1965), and the Metropolitan Achievement Test (MAT) (1965). On the WRAT Susie scored 4.4 in reading, 4.5 in spelling, and 4.7 in arithmetic. On the MAT (Elementary Battery) she scored 4.0 in word knowledge, 4.0 in word discrimination, 2.4 in reading, 4.0 in spelling, 4.0 in language, 4.0 in arithmetic computation, and 3.7 in arithmetic problem solving.

Susie was a student in the curriculum research classroom. In this class were six children, ages 8-12. Three of the children were referred from special classes, three from regular. The class was directed by two full-time teachers.

Materials

Three classes of subtraction problems were used throughout the project. Class 1 (C1) problems were the borrow facts where the top numerals or the minuends ranged from 10 to 19, the bottom numerals or the subtrahends ranged from 0 to 9; the remainders ranged from 1 to 9. There were 45 different problems in this class. Class 2 (C2) problems differed in that the minuends were from 20 to 98. The subtrahends were single digits, 1 through 9, and the remainders ranged from 11 to 89. In all of these problems, regrouping or borrowing was required. There are 360 problems in this class. For the Class 3 (C3) problems the minuends ranged from 20 to 98. The subtrahends ranged from 11 to 69; whereas the remainders were from 11 to 79. Regrouping was always required. There were 1260 different problems in this class.

From each class of problems five different sheets were constructed. For the C1 set some problems were used on more than one sheet since there were only 45 problems in the class. The problems were randomly assigned to the sheet and no problem appeared more than once on any sheet. For C2 and C3 different problems were placed on the five sheets in each class; no problem was used more than once. First, the 125 problems of each class were randomly selected from the total number of problems in each class. Then those problems were randomly assigned to one of the five daily sheets.

Design

The design of this study was a multiple baseline. Six phases comprised the study. During the first phase no contingencies were in in effect. Following this baseline, a withdraw contingency (withdraw of positive reinforcement) was successively scheduled on each sheet. This contingency, which specified that for each incorrect answer one minute of recess time would be lost, was arranged first on C1. Then, after five days, it was also placed on C2. After another ten days the contingency was additionally placed on C3. The contingency remained in effect for all three sheets for another twelve days.

The contingency was first removed from the C2 sheet. After three days, the contingency was removed from the C1 sheet. During the final phase the contingency was in effect for only the C3 sheet. Since the quarter ended, no time was left to schedule a condition where the contingency was removed from that sheet.

Procedures

Each day the pupil was given three sheets of subtraction problems. Each sheet contained 25 problems of a different class. On successive days a different arrangement for each class of problems was used. The order in which the sheets were given the pupil varied from day to day.

The pupil was requested to complete each page. Following the completion of the three pages, they were collected by the manager who checked the answers, counted the number of correct and incorrect responses, calculated the rate and percentage scores, then plotted the data.

Throughout this study the manager used scoring sheets to record information regarding each session. On these sheets, one for each class of problems, were recorded the day of the session, the date, number of correct and incorrect answers, time required to complete each sheet, correct and error rates, and the percentage correct score. Since each sheet contained 25 problems, each problem was worth 4%.

After the study was completed, all the daily papers were rechecked to ensure that the girl's responses were accurately graded. Each entry on the scoring sheets was also re-examined. The data appearing on the graphs were based on the rechecks.

Throughout the study the pupil was given no feedback as to which problems were incorrect. The reason no feedback was given and that no teaching was provided was that the manager had observed, on occasions prior to this study, that the pupil could perform the C1 and C2 problems. The manager did not, however, know what to expect when Susie worked the C3 problems.

During the phases when loss of recess time was scheduled for errors, she was simply told how many problems were incorrectly answered, and how much recess time she lost. Also, during these conditions, she was told the sheet(s) where the contingency applied. She was not told, however, when the contingency was removed.

RESULTS

The data from the three sheets; C1 (see Figure 1), C2 (see Figure 2), and C3 (see Figure 3) will be discussed by phases.

Phase 1: Baseline

During the baseline phase Susie's performance was very poor on all problems. On the C1 problems only one 100% day was noted; none were recorded for the C2 problems. On four of the five days on the C3 problems her scores were 0%.

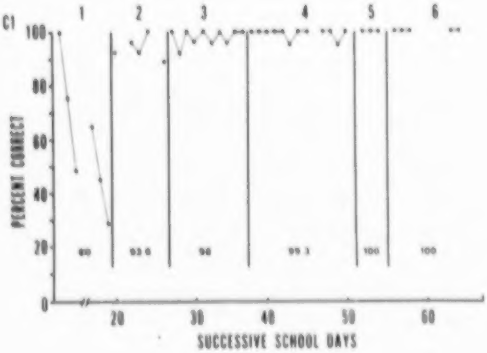


Fig. 1. Percent correct for C1 type subtraction problems throughout PROJECT 1. Only successive data days are connected by lines. The spaces between data days are accounted for by absences, weekends, or holidays.

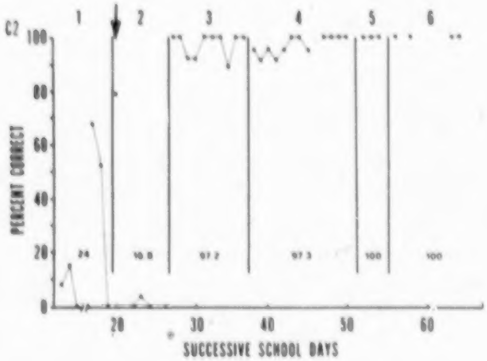


Fig. 2. Percent correct for C2 type subtraction problems throughout PROJECT 1.

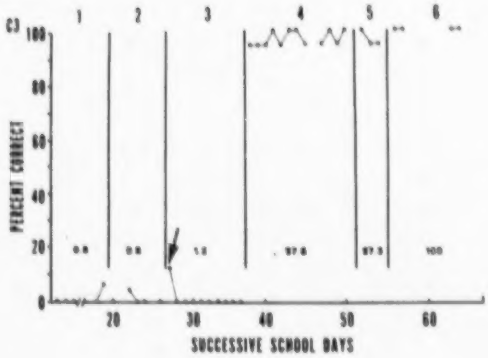


Fig. 3. Percent correct for C3 type subtraction problems throughout PROJECT 1. The arrow on the first day of Phase 3 indicates that she got the first three problems correct, but erred on all the rest that day.

Phase 2: Loss of Recess Time for C1 Errors

When the contingency of one minute taken from recess for each problem missed was placed on the C1 problems, the pupil's performance was immediately affected. Her score on the C1 problems rose from a mean of 60% in the baseline condition to a mean of 93.6% in Phase 2. Susie's performance on the C2 problems was positively affected only on the first day of Phase 2 as indicated by the arrow on Figure 2. After that, she again did poorly on those problems. On the C3 problems her accuracy continued to be quite poor.

Phase 3: Loss of Recess Time for C1 and C2 Errors

In the third condition the withdraw contingency was still in effect for the C1 problems. She continued to maintain a high percentage score (mean = 98%) for those items. During this phase the contingency was also arranged for the C2 problems. Her performance on those items was immediately and positively influenced. On the C2 items her mean percent score in Phase 2 was 16.8%; in the third condition, it rose to a mean of 97.2%. On the first day of this phase, she showed some improvement on the C3 problems as pointed out by the arrow on Figure 3. On that day the first three problems on that page were done correctly. This indicated for the first time that she could do at least some of the problems in that class. After the first day, however, her percent score remained at zero.

Phase 4: Loss of Recess Time for C1, C2, and C3 Errors

During this condition the withdraw contingency was in effect for all the problems. On the C1 problems Susie maintained a high mean of 99.3%. Her scores on the C2 problems also remained high, with a mean of 97.3%. Susie showed the same dramatic improvement on the C3 problems that she had on the two previous classes once the error contingency was applied. Her percent correct score rose from 1.2% in the previous condition to 97.6% in this phase.

Phase 5: Loss of Recess Time for C1 and C3 Errors

The contingency was still in effect for C1 and C3 problems but removed from C2. Susie maintained a mean of 100% on the C1 problems and 97.3% on C3. She had a mean of 100% on the C2 problems even without the contingency in effect.

Phase 6: Loss of Recess Time for C3 Errors

The experimenters now removed the contingency from the C1 problems. She maintained 100% on this class as well as on the C2 problems. Her scores on the C3 problems rose to an average of 100%.

DISCUSSION

When the withdraw contingency was imposed first on C1, then C2, then C3, the effects were immediate. When placed on C1 in the second condition, her percent score rose from a 28% score on the last session of the preceding phase to a 92% score on the first day of the withdraw condition. Similarly, her performance on C2, when the contingency was scheduled, rose from 0% to 100%, and on C3 from 0% to 96%.

These prompt changes, however, were noted on only the sheet where the contingency was scheduled, generalizations to the other sheets did not occur. In Phase 2 when C1 performance was influenced, the pupil's scores on C2 and C3 remained as low as they had been during the baseline. When the contingency was arranged for C1 and C2 in Phase 3, her performance on C3 continued to be poor.

A further observation, based on these data, was that performance on C1 and C2 continued at a high level when the withdraw contingency was removed. Thus, generalization of another type had occurred; when the treatment was removed, performance continued to be satisfactory.

*PROJECT II

The same classes of subtraction problems used in PROJECT I were used in this project. PROJECT II, however, differs from the first, both in design and tactic. The design of this project followed an ABA format; whereas the first project utilized a multiple baseline design. Instead of using contingencies to increase academic performance, teaching routines were employed in this project.

Beyond the primary purposes of this project--to study the effectiveness of some commonly used instructional aids--a secondary purpose was to obtain data about within and across response class generalizations.

METHOD

Subject and Setting

The subject of this study was a 10-year-old boy, Greg, referred to the EEU from a local special education class because of academic deficiencies. Greg was a member of the same class as Susie, the subject in PROJECT I.

When he entered the class, Greg was given two achievement tests. On the MAT (Elementary Battery) the following scores were obtained: 2.2 in word knowledge, 2.4 in word discrimination, 2.6 in reading, 2.2 in spelling, 1.2 in language, 3.0 in arithmetic computation and 2.3 in arithmetic problem solving concepts. On the WRAT, the child scored 2.5 in reading, 1.6 in spelling and 2.8 in arithmetic.

Materials

Greg received the same classes of problems used in the first project, but instead of computing one page of problems per class, he received two pages per class.

For each class, different problems were randomly assigned to one subclass or the other (i.e., C1*a* and C1*b*). Five different sheets for each subclass were devised. The order of presentation was the same as that described earlier.

Table 1 shows the total number of problems in each class and the number actually used in each subclass. Also indicated is how the problems were distributed on the five sheets.

TABLE 1

ORGANIZATION OF SUBTRACTION PROBLEMS IN PROJECT II.

CLASS	TOTAL	SUBCLASS	USED	5 SHEETS
C1	45	C1a	23	Modified Arrangement 20 per Sheet
		C1b	22	
C2	360	C2a	125	Different Problems 25 per Sheet
		C2b	125	
C3	1260	C3a	125	Different Problems 25 per Sheet
		C3b	125	

Design

The experimental strategy was, first, obtain baseline measurement from five sheets (C1*a*, C1*b*, C2*a*, C2*b*, C3*a*); then schedule a teaching aid with one sheet of the class currently being taught (e.g., C1*a*). Since one purpose was to measure within class generalization, the aid was not associated with the other sheet of the same class, (e.g., C1*b*). Another purpose was to obtain measurements of the "next" response class, to determine the extent of across class generalization (e.g., 2*a* and 2*b*).

Once criteria was reached on the sheet where the teaching aid was used, the aid would be removed to determine if performance continued at a satisfactory level. If performance maintained, the next step was to focus attention on the next class of problems (e.g., C2*a*). For that class a different teaching aid would be used.

Another aspect of the strategy was that after criteria on a class of problems was reached and instruction was shifted to a different class, measurement would continue to be gathered periodically on the former class of problems. By so doing, the experimenters would obtain data about retention and would be able to refocus teaching on those problems if the pupil's performance worsened.

It is important to remember that throughout this project the *a* sheets were those on which teaching was focused. The *b* sheets were the companion problems in each class; no teaching was directed toward them.

Procedures

Each day during the eleven conditions of this study, Greg was requested to complete each page of mathematics problems presented to him. When the student completed the sheets, the manager corrected each one, counted the number of correct and incorrect answers, calculated rate and percentage scores, then entered these on the scoring sheets and plotted the data. After the completion of the project, all the math sheets and the scoring sheets were rechecked. The data presented here were based on those rechecks.

PHASE 1: NO INSTRUCTION AIDS.

During the baseline condition no instructions, confirmation, or contingencies were scheduled. Measurements were obtained for all sheets except C3*b*, which was introduced later.

PHASE 2: PAPER CLIPS FOR C1a.

Throughout the second condition the pupil was given 18 paper clips. Each day prior to the boy's working the C1*a* problems, the manager handed him the clips and gave him one example. Greg was shown how, if the problem was 18 - 9 = ☐, to take

9 clips away from the pile, then count the remainder which was the answer.

PHASE 3: REMOVAL OF PAPER CLIPS FROM Cla.

During this condition, the paper clips were removed from the C1*a* problems. Greg received no instruction on any of the math problems in this phase.

PHASE 4: ABACUS FOR C2a.

In this condition, Greg was provided with an abacus as he worked the C2*a* problems. Each day during this phase the manager worked one problem for him.

PHASE 5: REMOVAL OF ABACUS FROM C2a.

The abacus was removed from the C2*a* problems. Greg received no instruction on any of the math pages during this phase.

PHASE 6: INSTRUCTIONS FOR Cla.

Attention was refocused on C1*a* because the boy's accuracy began to decline during the previous condition. Each day before Greg worked the C1*a* problems, instructions were provided. First, the manager wrote out a problem like 18 - 9 = ☐, then asked the pupil to say the answer. If he answered correctly, the manager said "good;" if not, she told him the answer. Second, Greg was asked if he needed to cross out anything (to regroup). If he answered correctly, the manager said "good." If he said that borrowing was required, it was explained why it was not necessary.

PHASE 7: REMOVAL OF INSTRUCTIONS FROM Cla.

Instructions were no longer given to the child for C1*a* during this phase. Greg received no instruction on any of the math problems during this phase. Measurement of his performance on the C3*b* problems began in this condition.

PHASE 8: CUISINARE RODS FOR C3a.

During this condition instruction was directed toward the C3*a* problems. Throughout this phase as the boy worked on the C3*a* problems, he was given a set of cuisinare rods to use. He was instructed in the use of the rods and given one example each day.

PHASE 9: REMOVAL OF CUISINARE RODS FROM C3a; ABACUS PROVIDED FOR C2a.

The cuisinare rods were not as effective as the paper clips or abacus. Also, during the eighth condition, Greg's accuracy on the C2*a* problems began to lapse. Attention, was, therefore, refocused on that class. During Phase 9 the abacus was again provided Greg as he worked the C2*a* problems.

PHASE 10: REMOVAL OF ABACUS FROM C2a.

The abacus was withdrawn from C2*a* in this phase. During this condition Greg received no instruction on any of the math problems. Greg did not, however, maintain a high level of accuracy.

PHASE 11: INSTRUCTION FOR C2a.

In the final phase, the focus was again on C2*a*. Now, instructions and an example were provided daily, prior to his answering those problems. First, the manager wrote out a problem of the type 24 - 6 = ☐. Second, she asked the pupil to solve it. If he was successful, she said, "good;" if not, the problem was computed for him. Third, the pupil was asked to explain the regrouping process (cross out 2, replace it with 1, place 1 by 4). If he could comply, the manager said, "good;" if not, the steps were explained.

Further conditions would have been run, but the school term ended.

RESULTS

Results for the classes and subclasses of PROJECT II will be discussed by phases. Graphs are included for C1*a* and C1*b* (see Figure 4), C2*a* and C2*b* (see Figure 5) and C3*a* and C3*b* (see Figure 6).

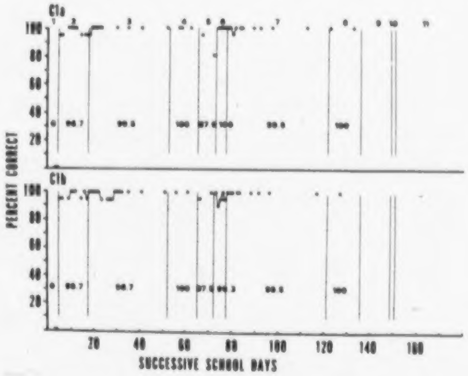


Fig. 4. Percent correct for C1a and C1b type subtraction problems throughout PROJECT II. Since the pupil's performance was extremely accurate during Phase 7, only review days were scheduled in Phase 8. No data were obtained throughout the final three conditions.

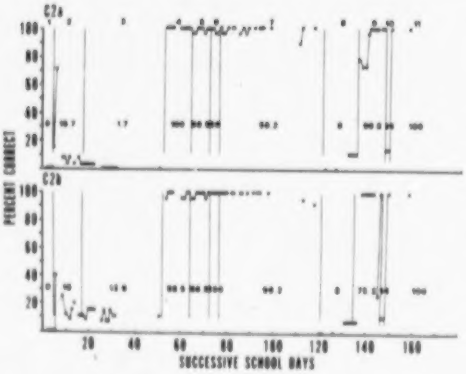


Fig. 5. Percent correct for C2a and C2b type subtraction problems throughout PROJECT II.

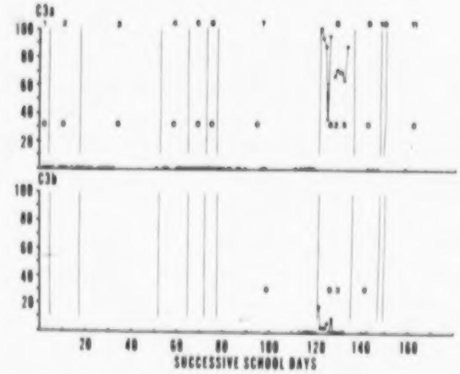


Fig. 6. Percent correct for C3a and C3b type subtraction problems throughout PROJECT II. No data were obtained on the C3b problems until the latter part of Phase 7 because the pupil erred on every problem on the C3a sheets.

Phase 1: No Instructional Aids

During the baseline condition when Greg received no instructions or any other interventions on the mathematics sheets, percent correct scores were zero on all five sheets. It is important to note that many of Paul's errors were not random. He used a faulty rule that accounted for many of his errors. He subtracted the smaller number in the units column from the larger number. His answer, for example, to the problem 18 - 9 = ☐ would be 1.

Phase 2: Paper Clips for C1a

When the paper clips were provided for C1*a*, his performance on those problems improved greatly ($\bar{x}$ = 96.7%). His performance was also excellent on the C1*b* problems; within class generalization had occurred. Meanwhile, on the C2*a* problems, his performance improved slightly from a mean of 0% in the baseline phase, to a mean of 15.7% in Phase 2. His scores ranged from 4% to 72%. On C2*b* he obtained a mean of 16%. Across class generalization was very minimal. His mean on C3*a* was zero.

Phase 3: Removal of Paper Clips from C1a

Although the paper clips were no longer provided for C1*a*, his performance continued to be excellent on that class with a mean of 99.5% for C1*a* and 98.7% for C1*b*. His scores on C2*a* declined from the previous phase. On 8 of 14 days his percentage scores were 0%. On C2*b* his scores ranged from 8% to 16%, a mean of 12.6%. All scores on C3*a* in this condition continued to be zero.

Phase 4: Abacus for C2a

In this condition an abacus was provided Greg as he worked on the C2*a* problems. His scores on those problems were all 100%. On the companion sheet, C2*b*, his scores were generally perfect ($\bar{x}$ = 98.5%). Within class generalization again had occurred. C3*a* scores were all zero; again across class generalization had not occurred. His average scores on the C1*a* and C1*b* problems throughout this phase continued to be 100%.

Phase 5: Removal of Abacus from C2a

The abacus was now removed from the C2*a* problems. During this time his accuracy on the C2*a* problems maintained. On 5 of the 8 days his scores were 100%. His accuracy, however, on the C1*a* problems began to worsen. On the two days during this phase that his performance was measured on C1*a*, his scores were 95% and 80%. On C1*b* the scores were 95% and 100%. All C3*a* scores were zero.

Phase 6: Instructions for C1a

Throughout the sixth condition, Greg received instructions on the C1*a* problems. His accuracy immediately improved on these problems; his four C1*a* scores were all 100%. On C1*b* no more than two problems were incorrect per day. His scores for C2*a* and C2*b* were all high. Once more, all C3*a* scores were zero.

Phase 7: Removal of Instructions from C1a

When instructions were removed from the C1*a* problems, he maintained high percentage scores. All but one of his C1*a* scores were 100% throughout this phase. All C1*b* scores were 100%. On C2*b* he had a mean of 98.2%; for C2*b* a mean of 99.2%. Throughout this phase all scores on the C3*a* problems were zero.

Phase 8: Cuisinare Rods for C3a

Greg was now given a set of cuisinare rods to use as he worked the C3*a* problems. On the first day of this phase his C3*a* score was 100%; thereafter, his accuracy was variable and generally declined. His scores ranged from 36% to 100% with a mean of 62.5%. On the C3*b* problems his scores ranged from 0% to 24% ($\overline{x}$ = 9.3%). Slight within class generalization occurred. His two scores on C1*a* and one score on C1*b* were 100% in this phase. Meanwhile, on C2*a* and C2*b* his scores were drastically reduced from the previous phase. For the first two days of Phase 8 his scores were zero on both sheets. For the next four days he got only three problems correct on each sheet.

Performance on C2*a* worsened in this phase because Greg had devised a new adaptation of his old rule. Now he reduced the tens column by 1, then subtracted the top number in the units column from the bottom number (e.g., 52 - 24 = 12). The focus throughout the next condition, therefore, was again on C2*a*.

Phase 9: Removal of Cuisinare Rods from C3a; Abacus Provided for C2a

The rods were withdrawn from C3*a* in this phase and the abacus was once again provided for C2*a* problems. His percent scores on C2*a* steadily increased in this phase; on the final six sessions his scores were all 100%. Although his performance on the C2*b* problems was poor on three days in this phase, his scores were 100% on 7 of 10 days. Once again within class generalization was demonstrated. Measurements were not taken on C1*a* or C1*b* during this phase. On C3*a* and C3*b* his scores returned again to zero.

Phase 10: Removal of Abacus from C2a

During this condition the abacus was once again removed from C2*a*. This condition ran only two days since the pupil's accuracy was very poor ($\bar{x}$ of 16%). Only four C2*a* and C2*b* problems were answered correctly on both days. No measurement was obtained for the C1*a*, C1*b*, or C3*a*, C3*b* problems in this condition.

Phase 11: Instructions for C2a

In the final phase, instructions were provided prior to Greg's working the C2*a* problems. During that brief phase, his scores on C2*a* and C2*b* were all 100%. Again, within class generalization occurred. Only one C3*a* measurement was obtained; that score was zero.

DISCUSSION

In several instances within class generalization was noted. In regard to Cl*a* and Cl*b* when, in the second condition, attention was directed toward the Cl*a* problems, not only did performance on that sheet improve, but his accuracy on Cl*b* problems was also aided.

Within class generalization was also noted when the focus was on C2*a*. In Phase 4, when the aid was given while working the C2*a* problems, not only was the accuracy on those items positively influenced, but his performance on C2*b* similarly improved. Within class generalization between C2*a* and C2*b* was again noted in Phase 9 when the abacus was, for the second time, offered the lad while he worked on the C2*a* problems. In Phase 11 such generalization was noted for a third time.

A modest amount of within class generalization occurred in Phase 8. When attention was focused on C3*a*, some slight improvement was noted on the C3*b* items.

Meanwhile, little, if any, across class generalization was observed. As the pupil's performance on Cl*a* and Cl*b* improved in Phase 2, his performance on C2*a* and C2*b* was only slightly influenced. No change was noted on C3*a*, the third class of problems. Similarly, in Phases 4, 9, and 11 when performance on C2*a* and C2*b* was influenced, accuracy on the C3 class was unaltered.

Throughout the experiment, three teaching aids were used. In Phase 2 paper clips were used on the Cl*a* problems; in Phases 4 and 9 an abacus was given the pupil as he worked on the C2*a* problems. Cuisinare rods were given the lad in Phase 8 as he worked on the C[illegible] problems. In every case effects were noted.

It should also be pointed out that in two instances when the aid was removed, in a phase following the involvement of the instructional device, his performance continued to be excellent. In Phase 3 when the paper clips were removed from Cl*a*, in Phase 5 when the abacus was withdrawn from C2*a*, his performance held up.

Performance, however, did not hold up when the abacus was removed from the C2*a* problems in Phase 10. Also, performance did not hold for the C3*a* problems when the rods were pulled off in Phase 9.

In addition to the three teaching aids that were used, instructions were scheduled for two sets of problems. In Phase 6 instruction was provided for the Cl*a* problems; instructions were also used in Phase 11 for the C2*a* problems. In both situations the effects were positive. Only in the first instance was the instruction removed in a subsequent phase. In Phase 7, when instruction was no longer used for the Cl*a* problems, his accuracy maintained at the level achieved in the previous phase. Since a quarter ended during Phase II, when instruction was arranged for the C2*a* problems, there was no opportunity to withdraw the instructions, hence evaluate their effects on those problems.

GENERAL DISCUSSION

A primary objective of this research was to point out that two remediation strategies are available for altering behavior; more specifically, the behavior of subtraction--reinforcement contingencies and teaching aids. Teachers should develop expertness in scheduling both; for if they rely on one set of events more than the other, they are probably less effective than they could be.

The professional teacher is the one who can discriminate student performance to the extent that contingent or non-contingent events are most appropriately scheduled. This research suggests some pupil behaviors to attend to in making the decision whether to use one type of procedure or the other, whether to establish a non-contingent or contingent event.

First of all, a careful error analysis is required before an effective remediation program can be established. A mere counting of the responses, correct and incorrect, will not generally inform the manager which instructional technique he should enlist. In the first experiment the girl's errors followed no apparent order, except that occasionally her "answers" were a series of numerals in consecutive order (e.g., 22, 23, 24). It was also noted that on occasion she erred on none of the problems. The first observation, that sometimes she simply wrote numerals in order, indicated that she was not making mistakes because of some erroneous rule. The second observation, that at times she could be accurate, signified that she could be precise when it appeared profitable to so behave. On the basis of these observations, it was hypothesized she would be amenable to contingency therapy. As for the selection of a reinforcer and a contingency, it was decided to use recess time, since earlier in the year she had been required to earn recess by her performance in reading and this had proved effective. In PROJECT I minutes were taken away contingent on errors, whereas in the reading study they were given contingent on correct answers.

In PROJECT II an error analysis in Phase I revealed that the boy subtracted the top numeral in the units column from the bottom numeral--a type of mistake commonly made by children. For example, his answer to 13 - 5 was 2, and invariably his answer was 6 to the problem 12 - 8.

Unlike the girl in PROJECT I, this pupil's mistakes were predictable. He had his own rule, which although consistently applied, was a faulty code. Other studies had been conducted with this lad where non-contingent events such as simple instructions were effective in altering his performance. In a reading study, for example, when he was told simply to "read a little faster," he generally did. This boy then seemed to be a good candidate for some non-contingent rehabilitation.

In conclusion, teachers must keenly and consistently observe and analyze the behavioral patterns of children in order to prescribe the most effective instructional technique. Teachers must be cognizant of the proficiency and quality of the pupil's performance; his error patterns that could be the result of erroneous rules; and, of course, his motivational system (which events or circumstances are reinforcing to him).

Often, if the diagnosis focuses on exact components of a pupil's behavior, the designing of an instructional technique is, if not obvious, rather simple. Many times, if the diagnosis has consisted of a systematic appraisal of the important elements of pupil behavior, the pupil will "tell" the manager which instructional plan to follow. Some aspect of his performance will clearly inform the teacher as to which tactic to schedule.

The teaching procedures used in these two projects were quite simple. In PROJECT I a rudimentary 1:1 withdraw contingency was arranged; in PROJECT II some uncomplex instructional aids were programmed. Such techniques can be easily used by any teacher. The teacher desiring to use the instructional events discussed here would merely need to be systematic in the involvement of the techniques, he would not need to read a complex instructional manual or attend an expensive workshop.

We, therefore, suggest that efforts be directed toward the simplification of all instructional components--designing rudimentary methods for diagnosis, devising uncomplicated procedures for instruction, developing basic routines for evaluation--instead of proposing highly complex and erudite educational systems. If such a frugal approach toward education were adopted, education would be more ecumenical; more people would be able to teach; hence, more individuals would learn.

REFERENCES

Becker, W. C., Madsen, C. H., Jr., Arnold, C. R. and Thomas, D. R. The contingent use of teacher attention and praise in reducing classroom behavior problems. *The Journal of Special Education*, 1967, 1, 287-307.

Jastak, J. R., Bijou, S. W. and Jastak, S. R. *Wide range achievement test*. (Rev. ed.) Wilmington, Delaware: Guidance Associates, 1965.

Lovitt, T. C. and Smith, J. O. Effects of instructions on an individual's verbal behavior. *Exceptional Children*, 1972, 38, 685-693.

Lovitt, T. C. and Curtiss, K. A. Effects of manipulating an antecedent event on mathematics response rate. *Journal of Applied Behavior Analysis*, 1968, 1, 329-333.

Metropolitan achievement test. Elementary Battery. New York: Harcourt, Brace and World, 1965.

Staats, A. W. and Butterfield, W. Treatment of nonreading in a culturally deprived juvenile delinquent: An application of reinforcement principles. *Child Development*, 1965, 36, 925-942.

⁂

SUPERIMPOSING CONTRACTS UPON A TOKEN ECONOMY

joseph r. jenkins
sheila gorrafa

The effects of several incentive systems on the reading and arithmetic performances of twelve primary EMR children were investigated. "Normal" (non systematic) classroom management was compared with a token economy plus superimposed contracts over an eight week period. Both reversal and multiple baseline designs were employed. Reading and arithmetic performances were highest when the token economy was combined with a superimposed contract. Performance was lowest when neither the token economy nor the combination of tokens and contracts were in effect.

Two approaches to contingency management in the classroom are contingency contracting (Homme, 1969) and the more general token systems (Neisworth, Deno and Jenkins, 1969). In contingency contracting an agreement is struck between two parties, the teacher and her students. The agreement in its general form states that when the student accomplishes a particular performance then some event, activity or object will be provided him by the teacher. Under a token system, on the other hand, the teacher sets a token value on a number of different student performances. At the same time she identifies a variety of events, activities or objects for which the tokens may be exchanged. These two approaches, of course, are closely related. In fact, contingency contracting as described here is a special case of what occurs in a token economy. The major difference is that under contingency contracting there is a one-to-one relation between the performance and its consequence which renders unnecessary a medium of exchange. In most instances either one or the other approach is employed separately. The study reported here investigates the effects of a token system and of a superimposed contingency contract on the academic performance of twelve EMR children. Data on the children's academic performance were collected under conditions of no tokens, tokens, and tokens plus a superimposed contingency contract.

METHOD

Subjects and Setting

The study was conducted at the Experimental Educational Environment (EEE), a Newark School District ESEA Title III Program which is located at the University of Delaware. Twelve EMR children whose ages range from 6 to 11 participated in the project.

The EEE is administered on a contingency management system with an ancillary token economy (poker chips) where children are rewarded for academic and socially appropriate performances. Work sheets and programmed texts are provided regularly. Individual contracts are established on a piecework basis in which there is a specific relation between the amount of work performed and the number of tokens earned. The classroom operates on a free spending system (Neisworth, Deno and Jenkins, 1969) in which the student chooses how and when he will spend his earnings. The child has the option of stopping work on an academic task after he has acquired sufficient tokens to enter the reinforcing events (RE) area, or, of remaining at work in the task area where he may continue to earn tokens for later spending. The RE area is located in the rear of the classroom. There, the children may engage in a number of high preference behaviors, such as bowling, painting, writing on the chalkboard, etc. A store within the RE area contains other commodities such as coloring books, bubble bath, candy, (the larger items are purchased on a lay-away plan), etc., which may be purchased at the end of the school day.

Recording and Reliability

The dependent measures in the study were pages completed by the children in their programmed readers (Sullivan Associates, 1969) and the number of correct problems completed on their daily arithmetic work sheets.

Each student kept a record sheet at his desk. This sheet had spaces for recording the page numbers from the programmed reader successfully completed and the number of arithmetic problems which had been worked correctly. The teacher and her aide acted as recorders, checking each child at his seat a minimum of two times in the morning and again in the afternoon immediately before the children left. A student was checked on the average of five times per day. At the end of each check the recorder paid the student for the number of pages and problems correctly completed. During these checks the student would present his reading answer sheets, where he had recorded his answers to each program frame, and also read upon request sample frames from each program page that he had completed since his last check. The recorder corrected the reading answer sheets. If there were more than three errors the student was asked to redo that page. If there were three or fewer errors, the recorder chose two frames from each page for the student to read orally. When a student misread a word the teacher pointed to the word and asked the student to "sound it out." If the student did not read the word correctly after this procedure then it was counted as an error. If more than two of these uncorrected errors occurred, the page was not counted in the student's total. Whenever frequent errors occurred the teacher or aide would spend a short time helping the child with the words which posed a difficulty for him.

When the child finished reading the sample frames the recorder marked on the student's record sheet those page numbers that the student had completed successfully and paid him the number of tokens he had earned.

Arithmetic problems were also recorded during these checks. The recorder corrected the completed worksheets on the spot, designating problems correct and wrong. The number of correct problems was marked on the record sheet and the student was paid for those. Students were always encouraged to correct their mistakes. On later recording checks the student was given credit for and paid for corrected problems.

An elevated observation deck and one-way mirror overlooked the classroom. Spot reliability checks on the reading and arithmetic measures were made throughout the year both from this observation deck and from within the classroom. For reading a reliability check was made by a judge other than the teacher or aide who listened as the child read sample frames to the recorder. An independent judgment was made for each page as to whether the child had performed satisfactorily (three or fewer errors on the answer sheet and two or fewer errors on the frame reading). Agreement between observer and recorder was consistently high, concurrence on over 90% of the pages checked.

For arithmetic recording, reliability was checked by periodically recorrecting arithmetic work sheets. Correction errors rarely exceeded 5%.

Treatment Sequence

A multiple baseline design was employed wherein treatments were applied to reading and mathematics independently.

The study lasted eight weeks. Week 1 was a normal token week. The token economy had been operative for seven months. During this and other token weeks which follow the children earned and spent tokens for performances and events which were described above. Week 2 was the first superimposed contract week. In this and subsequent superimposed contract weeks children continued to learn and spend tokens as they had in a regular token week. In addition, however, they negotiated an individual contract with the teacher which stated the amount of reading and arithmetic they must perform to earn an extra privilege. The teacher used as a guideline in her negotiations the child's high performance days during the token weeks. In Week 2 the superimposed contract involved outdoor play. For each child a contract was set which stated as soon as both a specified number of pages in his programmed readers and a specified number of arithmetic problems had been correctly completed he could play outside until 11:00 A.M.

The purpose of this and the subsequent superimposed contract week was to determine if reading performance could be increased, but not at the cost of decreasing arithmetic performance. Therefore, the contracts called for approximately the same number of arithmetic problems completed as in the previous week. Only the reading was augmented.

Week 3 was again a superimposed contract week. This time the children contracted for a sight-seeing train ride. Each child negotiated a combination reading and arithmetic contract on Monday for the trip which was planned for Thursday. Again only reading was augmented.

In Week 4 the class returned to tokens with no additional contracts, the same conditions which prevailed during Week 1.

Week 5 was a no token week. On Monday the teacher informed the class that the store was closed. She paid no tokens for performance that week, nor were play time and other high preference activities contingent on any student performance.

During Week 6 the token economy was resumed. The conditions of this week were the same as those during the Weeks 1 and 4.

Weeks 7 and 8 were both superimposed contract weeks. Monday of the seventh week was not a full school day. Tuesday of that week the teacher announced that a picnic in a nearby park was scheduled on Friday afternoon for those children who wished to contract for it. Again, individual contracts were written stating the number of programmed reader pages and arithmetic problems needed for admission to the picnic. On this and the subsequent week both reading and arithmetic performances required for fulfillment of the contract were increased relative to the children's performance in Week 6.

On Monday of the eighth week the teacher announced that on the subsequent Friday the class could earn a full day bus trip to the beach. Again, individual contracts were negotiated. Arithmetic contract demands were augmented while reading requirements were approximately equivalent to those of the previous week.

RESULTS

The mean performances of the students over eight weeks in reading and arithmetic are shown in Figures 1 and 2.

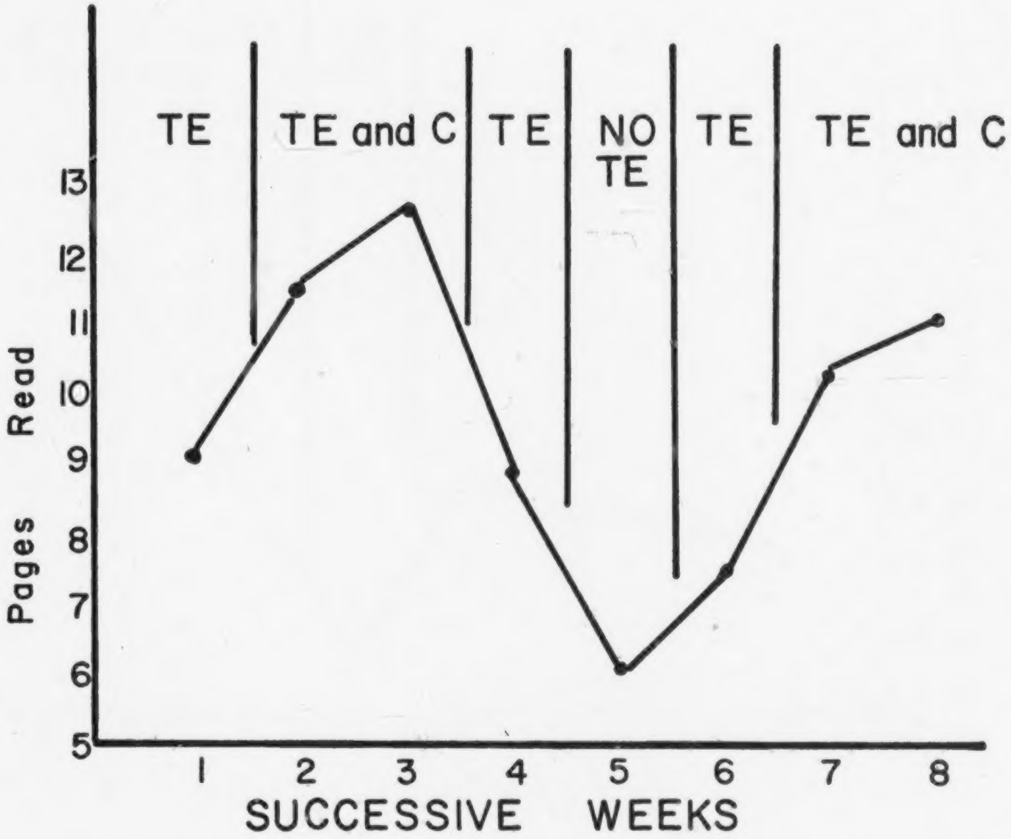


Fig. 1. Mean pages read/day by individuals during the Eight Weeks.

The mean performance of the class shifts appropriately with each change in contingencies. The highest performances are obtained when superimposed contracts are in effect. The normal token economy produces the next highest performance. Finally, the lowest performances are observed when neither the token economy nor the contract is operative.

Comparisons of individual children's performance in successive weeks was revealing.

Reading Performance

The reading performance of ten of the twelve children increased from Week 1 to Week 2, tokens to a superimposed contract, $p < .02$, sign test. In the transition from a superimposed contract to tokens, Week 3 to

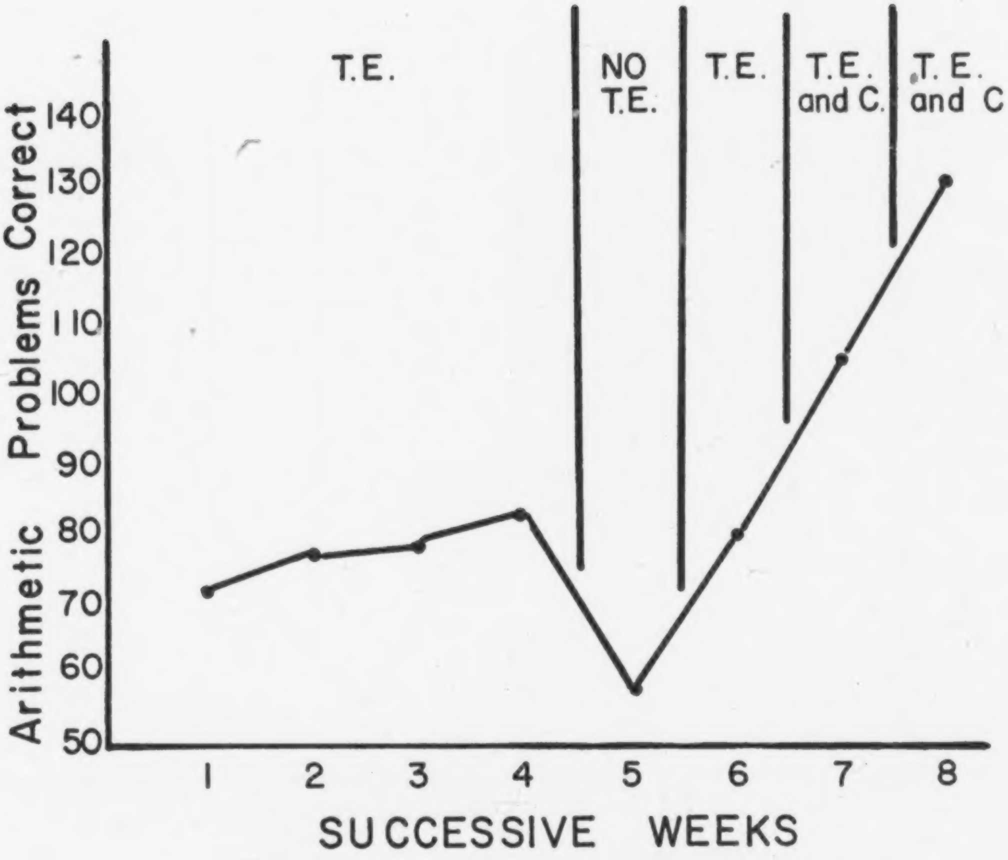


Fig. 2. Mean arithmetic problems completed/day by individuals during the Eight Weeks.

Week 4, eleven of the twelve children decreased their output, $p < .003$, sign test. The effect of terminating the token economy is observed in the decreased output of eleven of the twelve children from Weeks 4 to 5, $p < .003$, sign test. Resumption of the token economy was not particularly impressive in that only eight of the twelve children made increases. A gradual increase in performance occurred over the first three days of the renewed token economy (Monday, 5.17; Tuesday, 7.06; and Wednesday, 9.50).

Superimposing a contract again increased performance above that observed under the token system by itself, Week 6 to Week 7, with eleven of the twelve children showing an increase, $p < .003$.

Arithmetic Performance

Since the contracts for arithmetic demands were not augmented significantly during the first two superimposed contracts, the first interesting data on arithmetic performance involve the effects of terminating the token economy. With this change from Weeks 4 to 5, ten of the twelve children decreased their arithmetic production, $p < .02$, sign test. Subsequently in Week 6 when the token economy was reinstated, ten of the twelve children increased in arithmetic performance.

When a superimposed contract with an augmented arithmetic demand was instated, Week 7, all twelve of the children increased their arithmetic output, $p < .001$, sign test. Finally, in the eighth week, the last superimposed contract week, eleven of the twelve children completed more arithmetic problems than in the previous week ($p < .003$), consistent with the demands of their contracts.

The data suggest that token systems and contingency contracts can be combined to influence academic performance. Further investigation of this topic would be welcome, particularly in regards to criteria used to determine the conditions of contingency contracts. Determining the amount required of students for contract fulfillment is at times difficult for the practitioner. In the present study the teacher relied upon both her own intuition and upon the children's past performance.

There is another problem with using contracts exclusively, especially when the contract has a time constraint. If a child fails to complete his contracted reading and arithmetic he is left behind while the other children visit the beach or see a play. There is some security in general token systems in as much as time is generally not a factor and there is a variety of events differing in cost for which tokens may be exchanged. Seldom is a child left without some reinforcing alternative.

REFERENCES

Homme, L. *How to use contingency contracting in the classroom.* Urbana: Research Press, 1969.

Neisworth, J. T., Deno, S. L. and Jenkins, J. R. *Student motivation and classroom management.* Newark, Delaware: Behavior Technics Press, 1969.

Sullivan Associates Program. St. Louis: Webster Division, McGraw-Hill, 1968.

◆ ◆ ◆ ◆

part 7:

BEHAVIOR ANALYSIS IN higher education

COLLEGE STUDENT STUDY BEHAVIOR IN A PERSONALIZED INSTRUCTION COURSE AND IN A LECTURE COURSE

david g. born; michael davis
paul whelan; donald jackson

The procedures comprising the personalized system of instruction (PSI) described by Keller (1966, 1968) contrast sharply with traditional methods of college instruction. In comparison with lecture instructional formats, PSI has been shown to produce superior performance on traditional examinations (Born, Gledhill and Davis, 1972; McMichael and Corey, 1969; Sheppard and MacDermot, 1970), more favorable ratings on questionnaires (e.g., Born and Herbert, 1970; Gallup, 1969), and in some cases a higher percentage of withdrawals from courses (Born, 1971a, 1971b). The present experiment attempted to determine how the amount and distribution of student study over an academic term would differ under PSI and a lecture dominated instructional system.

METHOD

The amount of time spent studying by students enrolled in a PSI section and a lecture section of beginning psychology was monitored by making course materials available only in a special Study Center where a monitor checked out study units to students and noted (a) the time of entry, (b) the unit requested, and (c) the time of exit. A study unit included both the printed material comprising the reading assignment and a lengthy set of study items. Students in both sections had access to the same material presented in the same format.

PSI Procedures

Course material was packaged in a series of eighteen units and all students were required to pass a short written test over each unit of material. Each unit test was evaluated by a proctor, and students were allowed to progress to the next unit only if each item was judged correct. If the student missed one or more items he was asked to re-study and take a different examination over the same unit of material. Because there were approximately fifty percent more days scheduled for examinations than there were examinations to be taken, students had some opportunity to complete the course at their own rate. Twice a week (Tuesday & Thursday) students could attend a series of optional lectures, movies, and demonstrations which supplemented the reading material.

Lecture Course

Students enrolled in the lecture course could attend lectures closely related to the course reading material on Monday, Wednesday, and Friday, and could join students in the PSI section on Tuesday and Thursday for the special series of lectures, demonstrations, and films. Attendance at all class meetings was optional and students were told they would only be tested over the written material comprising the study units.

Major Examinations

Students in both the PSI and lecture classes took the same hourly midterm and final examination. Each of these exams consisted of multiple choice and short-answer essay items and was given at two different times, about one week apart, to permit students to select their own examination date. Students in the PSI class had to complete the first nine units of course material prior to taking the midterm examination over that material, and the last nine unit exams prior to taking the final. None of the exam items had been used earlier in the course as study questions or questions on unit examinations.

RESULTS AND DISCUSSION

An analysis of exam scores revealed that the superiority of the mean test performance of the PSI section was statistically reliable ($p < .05$); similar results have been reported by earlier investigators (e.g., McMichael and Corey, 1969). Further, there were large and statistically reliable differences in the mean total study time invested in the course by students in the two sections (see Figure 1). Relative to the mean of 30.34 hours of study in the lecture section, the mean of 45.62 hours of study in the PSI section represented a fifty percent increase in total study.

An analysis of mean cumulative study time prior to taking the midterm (see Figure 2), and from the midterm to the final examination, revealed a fairly constant average rate of study by early and late exam takers in both sections, but the slope of the mean cumulative graphs was quite different. However, there was substantial variation in both the pattern and total amount of study by individual students prior to each exam (see Figure 3). For some students (e.g., P3, P4, L5) study commenced early and occurred regularly up to an examination; for others (e.g., L1, L2) study started later in the term and/or it was erratic.

Although a greater total amount of study occurred in the PSI section, if one adds the average number of hourly lectures attended ($\bar{X} = 19$) to average study time for the lecture class, the total time of preparation for students in the two classes is nearly identical. Considering that PSI produces superior (learning) performance on examinations, relative to our lecture format, for comparable hours of preparation PSI would seem to be the method of choice for students.

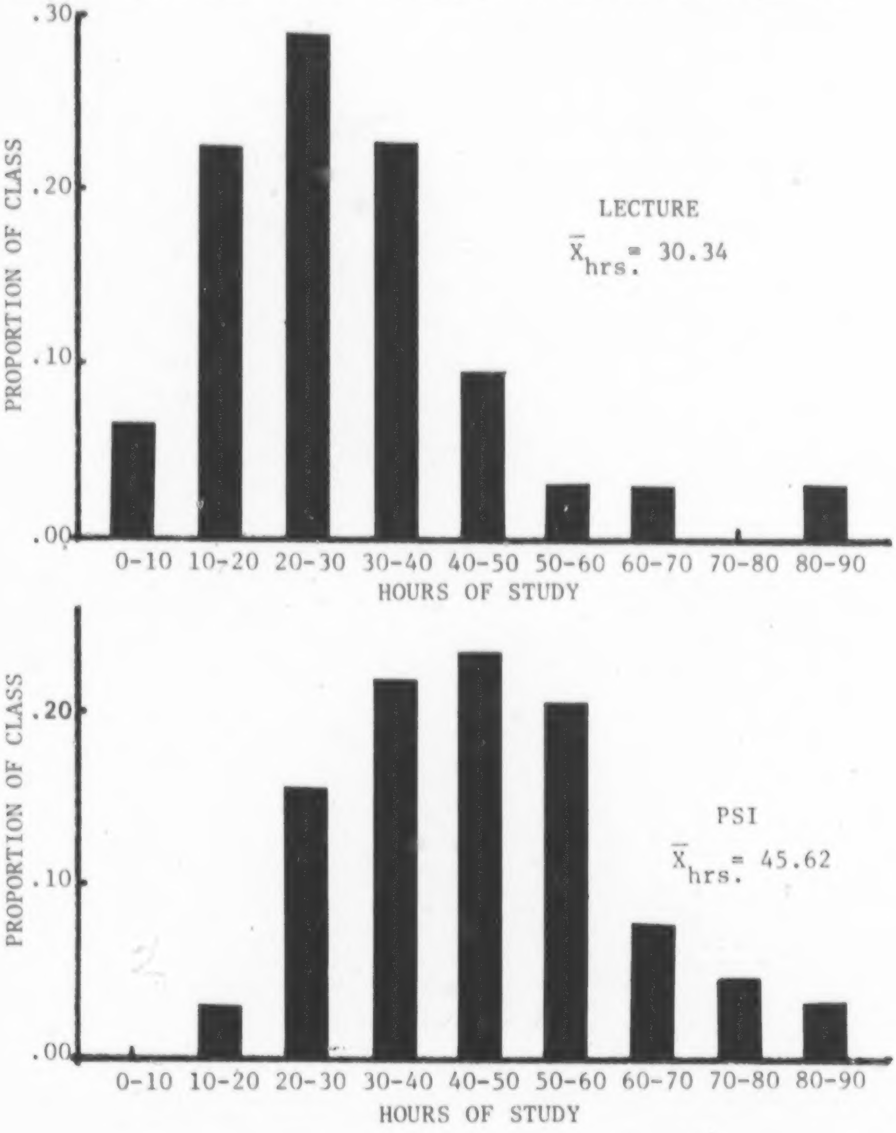


Fig. 1. Histograms showing total number of hours of study by students in the PSI and lecture section of a class in beginning psychology. Because there were different numbers of students in the two sections the absolute frequencies falling in each 10-hour interval have been converted into proportions for each section.

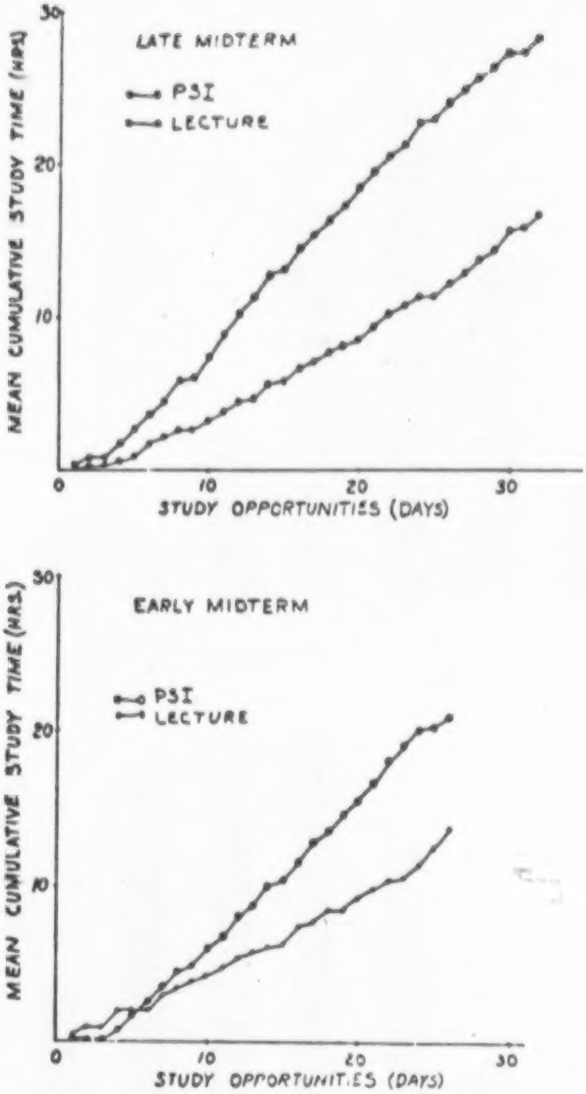


Fig. 2. Mean cumulative study time for students taking the early and late midterms in the PSI and lecture sections.

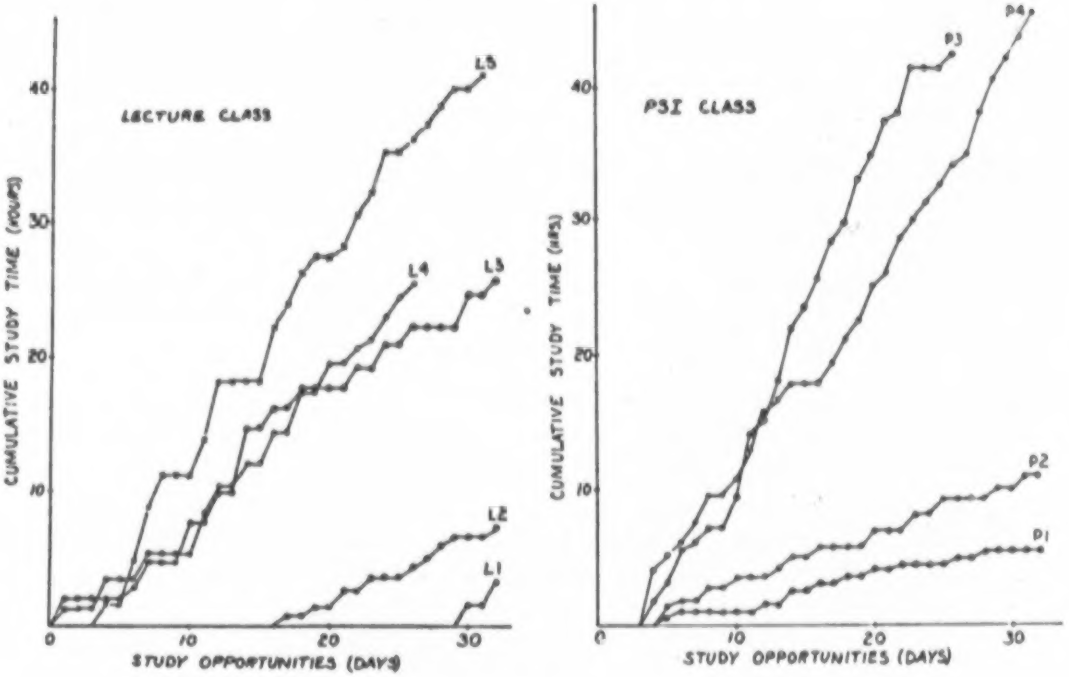


Fig. 3. Individual study profiles up to the midterm exam for several students in the PSI and lecture class. The students selected were the two students from each section who spent the least time in preparation (L1 and L2, P1 and P2) and the two students who spent the most time (L3 and L5, P3 and P4). Because L5 was a very atypical student in the lecture section, the record of L4 has also been included.

REFERENCES

Born, D. G. Student withdrawals in personalized instruction courses and in lecture courses. In F. Newman (Chm.), *Personalized instruction: A national trend moves into the Rocky Mountain region.* Symposium presented at the meeting of the Rocky Mountain Psychological Association, Denver, May, 1971.

Born, D. G. Effects of the Keller plan on students. Paper presented at the M.I.T. Conference on the Keller Plan, Cambridge, Massachusetts, October, 1971.

Born, D. G. and Herbert, Emily W. A further study of Keller's personalized system of instruction. *Journal of Experimental Education*, 1971, 40, 6-11.

Born, D. G., Gledhill, S. M. and Davis, M. L. Examination performance in lecture-discussion and personalized instruction courses. *Journal of Applied Behavior Analysis*, 1972, 5, 33-43.

Gallup, H. F. Personalized instruction in introductory psychology. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, Illinois, May, 1969.

Keller, Fred S. A personal course in psychology. In R. Ulrich, T. Stachnik and J. Mabry (Eds.), *Control of human behavior.* Glenview, Illinois: Scott Foresman, Inc., 1966, 91-93.

Keller, Fred S. "Goodbye, teacher. . ." *Journal of Applied Behavior Analysis*, 1968, 1, 79-89.

McMichael, J. S. and Corey, J. R. Contingency management in an introductory psychology course produces better learning. *Journal of Applied Behavior Analysis*, 1969, 2, 79-83.

Sheppard, W. C. and MacDermot, H. G. Design and evaluation of a programmed course in introductory psychology. *Journal of Applied Behavior Analysis*, 1970, 3, 5-11.

THE MEASUREMENT & ANALYSIS OF COLLEGE STUDENT STUDY BEHAVIOR

james m. johnston
michael d. roberts
george w. o'neill

The program of behavior analysis which Drs. Pennypacker and Johnston began in 1968 has as its goals the systematic empirical investigation of college level teaching techniques and student academic performance, and the resulting development of improved techniques which will produce improved performance. To date the study behavior of students has been among the variables least investigated by ourselves and by others in this area, and this lack of research emphasis seems to be unusually disproportionate to its potential importance. A student's study behaviors are probably the class of responses most influenced by teaching procedures and the most influential in affecting his academic performance. The reasons for this disproportionate emphasis most probably stem from the fact that in most institutions of higher education, student study behavior both within and across students occurs at the choice of the individual student in a wide variety of physical settings at all hours of the day and night. Thus, unless a teacher can control the time and place of his students' studying, there exist considerable problems of observation and measurement.

This difficulty dictated some form of self-reporting by the individual student in such a way that the measurements would be as direct, continuous, and accurate as possible. The first step was to construct a fairly exhaustive list of the possible variety of different study responses. This list was then organized into a concise format which was easy to understand and quick to use.

The first Study Report Form (SRF) which resulted is shown in Figure 1. The top of the sheet has places for identifying information and is followed by some brief instructions. After circling the total time studied for the period under measurement (to the nearest half hour), the student rated each separate study response he emitted on a 0 to 5 continuum. The ratings somewhat crudely denoted the amount or proportion of the total time circled which was devoted to each of the separate study responses rated. The ratings ranged from none of the time (0) to all of the time (5).

Both this form and a subsequent version which was experimented with suffered from at least one major disadvantage. The students completed the form only when they came in to perform on the course material (an average of twice a week), and they therefore had to summarize and report

STUDY REPORT FORM TS_____ STUDENT__________

Date_____Unit_____Try_____RC_____RIC_____Crit.C_____Crit.IC_____

Carefully complete this form describing how you prepared for this performance session by applying the following numbers and descriptions of how much of your total time you spent in the appropriate categories.

1=*very little of the time*, 2=*little of the time*, 3=*some of the time*, 4=*most of the time*, 5=*all of the time*

Total time spent preparing for this performance session (circle one):

1/2, 1, 1 1/2, 2, 2 1/2, 3, 3 1/2, 4, 4 1/2, 5, 5 1/2, 6, 6 1/2, 7, 7 1/2, 8, 8 1/2, 9, 9 1/2, 10, 10 1/2, 11, 11 1/2, 12 hours

★ENVIRONMENTAL CONDITIONS DURING PREPARATION
own room.. ______
library.. ______
other__________.. ______
overall level of disturbance low ___ 1 ___ 2 ___ 3 ___ 4 ___ 5 high

★READING (generated written aids? *yes*___ *no*___)
text... ______
supplementary materials.................................. ______

★REREADING (generated written aids? *yes*___ *no*___)
text... ______
supplementary materials.................................. ______

★WRITTEN AIDS
underlining, brackets, etc. .... ______
outlining.. ______
summarizing.. ______
key terms, concepts (definitions)........................ ______
writing in book.. ______
constructing: *fill-in questions*........................ ______
multiple choice questions.............................. ______
examples... ______

★SELF-QUIZZING
text: *oral*... ______
silent... ______
written.. ______
written aids: *oral*..................................... ______
silent... ______
written.. ______

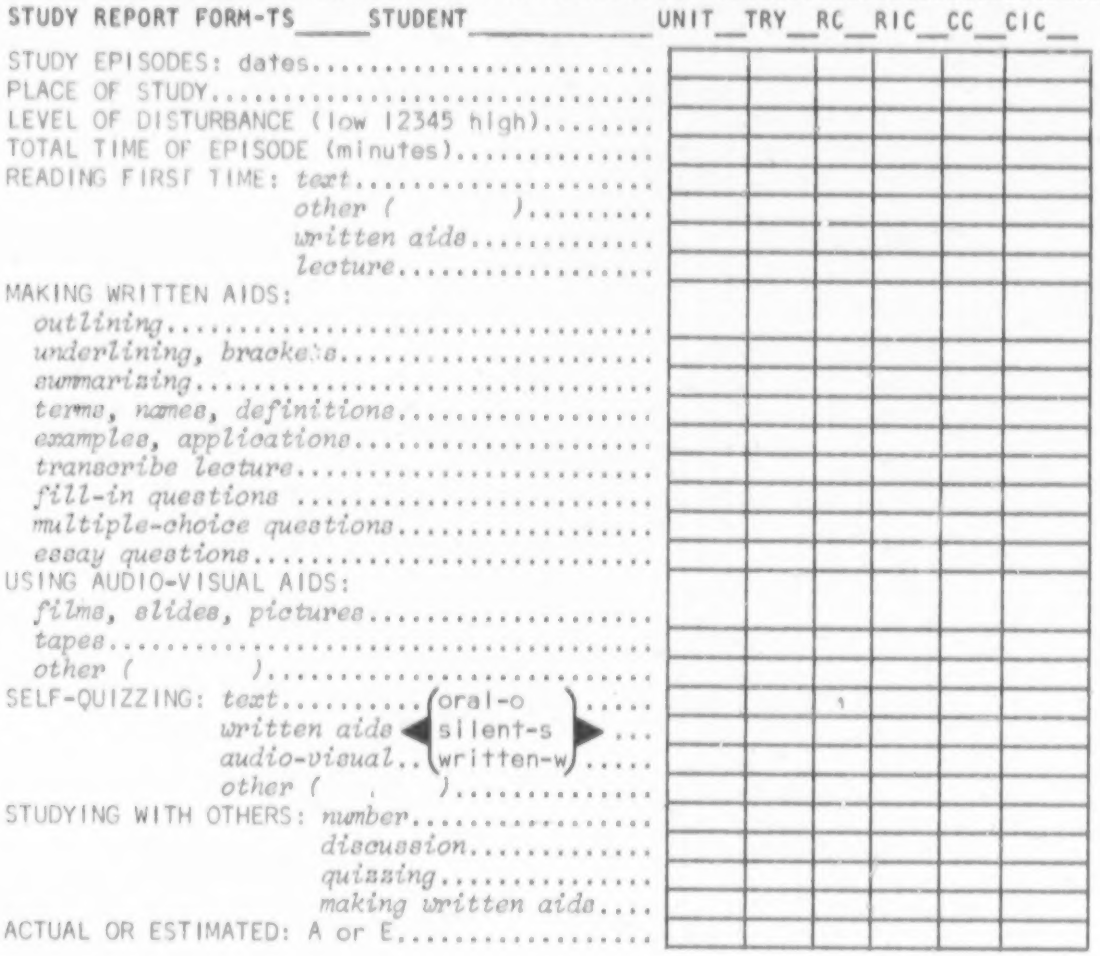
★WORKING WITH OTHERS (how many?____)
discussion... ______
quizzing... ______
exchanging written aids.................................. ______

How can we change this form so that it more accurately describes your preparations for performance sessions?

Fig. 1. Original Study Report Form.

from memory all of the various study episodes which occurred since the last time they were in to perform. This was the reason for requiring only the relatively crude ratings rather than a reporting of actual times for the separate study behaviors. It seemed that regardless of any increased precision which might be designed into the form itself, the delayed nature of its use guaranteed a certain minimum of inaccuracy and coarseness in the data.

This conclusion resulted in the most recent revision shown in Figure 2 which incorporated a number of significant changes. The separate pieces



STUDY REPORT FORM-TS____ STUDENT____________ UNIT__ TRY__ RC__ RIC__ CC__ CIC__

STUDY EPISODES: dates........................						
PLACE OF STUDY...............................						
LEVEL OF DISTURBANCE (low 12345 high)........						
TOTAL TIME OF EPISODE (minutes)..............						
READING FIRST TIME: *text*......................						
other ( ).........						
written aids.............						
lecture..................						
MAKING WRITTEN AIDS:						
outlining...................................						
underlining, brackets........................						
summarizing.................................						
terms, names, definitions...................						
examples, applications......................						
transcribe lecture..........................						
fill-in questions						
multiple-choice questions...................						
essay questions.............................						
USING AUDIO-VISUAL AIDS:						
films, slides, pictures.....................						
tapes.......................................						
other ( )...........................						
SELF-QUIZZING: *text*......... oral-o						
written aids ◀ silent-s ▶ ...						
audio-visual.. written-w						
other ( ).............						
STUDYING WITH OTHERS: *number*................						
discussion............						
quizzing..............						
making written aids....						
ACTUAL OR ESTIMATED: A or E.................						

Fig. 2. Second revision of the Study Report Form.

of information requested were edited and reorganized, making it possible to reduce the size of the form to 5" x 8" so that it could be kept in the student's textbook. This was done because now the student was instructed to fill out the SRF during or immediately after each individual episode of studying which occurred between scheduled performances. However, instead of approximating the time for each studying response by 0-5 ratings, he now recorded the actual time (to the nearest minute) he spent engaged in each kind of study response. This change permitted the possibility of a more detailed examination of the pattern of studying across separate episodes as well as considerably more direct and accurate observations. The reverse side of the SRF contained abbreviated instructions and explanations of the different studying responses (Figure 3).

INSTRUCTIONS FOR STUDY REPORT FORM

★*IMPORTANT:* Complete during or immediately after each study episode.

STUDY EPISODE - a relatively uninterrupted period of study separated from other episodes by other activities.

PLACE OF STUDY - write in where the episode occurred.

LEVEL OF DISTURBANCE - rate 12345 any visual, auditory, or other distractions.

TOTAL TIME OF EPISODE - mark actual time of entire episode to the nearest minute.

READING FIRST TIME - reading for the first time ever.

REREADING - any reading except the first reading.

MAKING WRITTEN AIDS:

- *Outlining* - constructing formal or informal outlines of any part of the material.
- *Underlining*, *Brackets* - marking any parts of the text, written aids, lecture, other.
- *Summarizing* - writing summaries of units, chapters, paragraphs, etc.
- *Terms*, *Names*, *Definitions* - writing terms, definitions, concepts, important points, etc.
- *Examples*, *Applications* - writing examples, applications, uses, etc.
- *Transcribe Lecture* - copying yours or other lecture notes or transcribing taped lectures.
- *Questions* - constructing any of the listed types of questions.

USING AUDIO-VISUAL AIDS:

- *Films*, *Slides*, *Pictures* - time spent viewing.
- *Tapes* - time spent listening.

SELF-QUIZZING - mark time spent in oral, silent, or written self-quizzing on any material.

STUDYING WITH OTHERS:

- *Number* - mark the number studied with.
- *Discussion* - talking about material with others.
- *Quizzing* - quizzing each other orally or written.
- *Making Written Aids* - based on group discussion or others' written aids.

ACTUAL or ESTIMATED - mark A or E if most times are actually correct or estimated later.

Fig. 3. Instructions for second revision of the Study Report Form.

Following the first semester's use of this version of the SRF, the students in an abnormal psychology course were asked to complete an anonymous questionnaire regarding their use of the forms. They rated on a 1 to 5 scale the extent of their disagreement or agreement with a series of statements. A 1 signified strong disagreement and 5 signified strong agreement. Some of their reactions are shown in Figures 4 through 9.

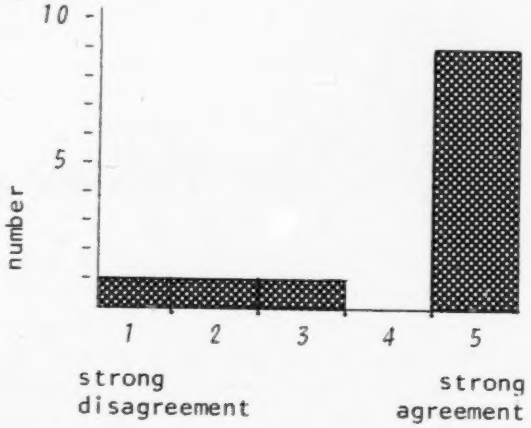


Fig. 4. Student responses on a 1 to 5 agreement scale to the statement, "I usually kept the SRF's in my text."

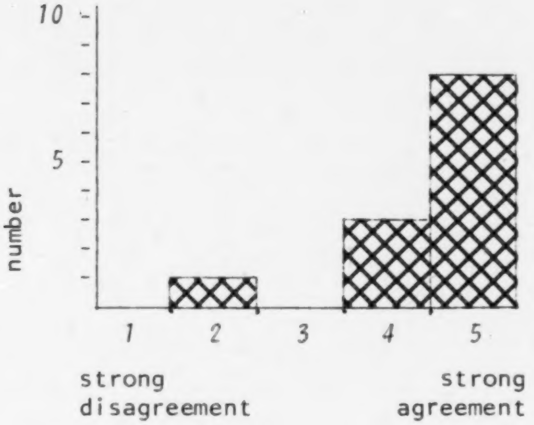


Fig. 5. Student responses on a *1* to *5* agreement scale to the statement, "I usually completed the SRF during or soon after each episode of study."

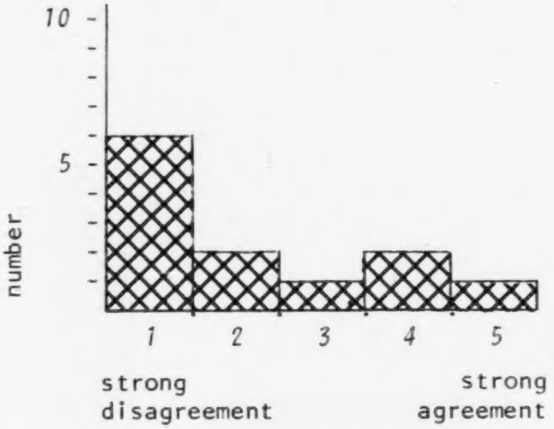


Fig. 6. Student responses on a *1* to *5* agreement scale to the statement, "I usually estimated my times rather than using a watch or a clock."

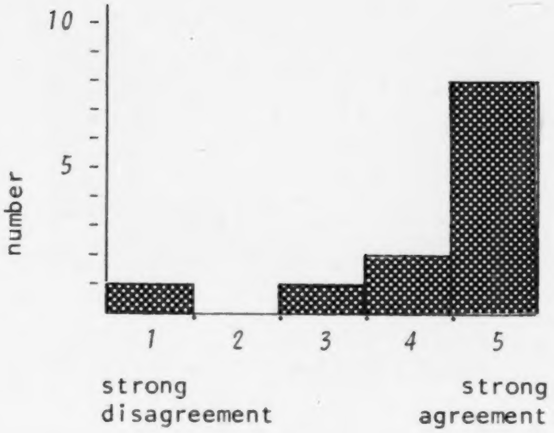


Fig. 7. Student responses on a *1* to *5* agreement scale to the statement, "After I got used to it, the SRF was easy to fill out."

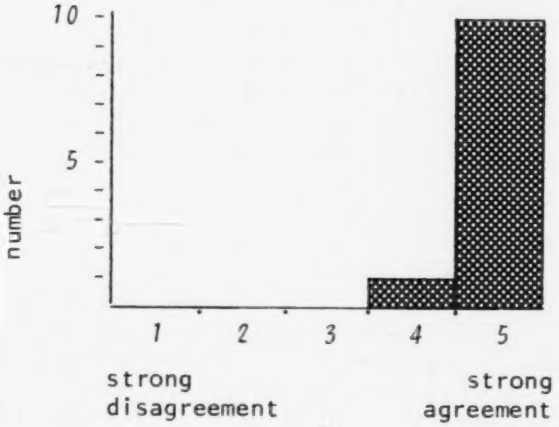


Fig. 8. Student responses on a *1* to *5* agreement scale to the statement, "I really didn't mind if my Student/Teacher saw my SRF."

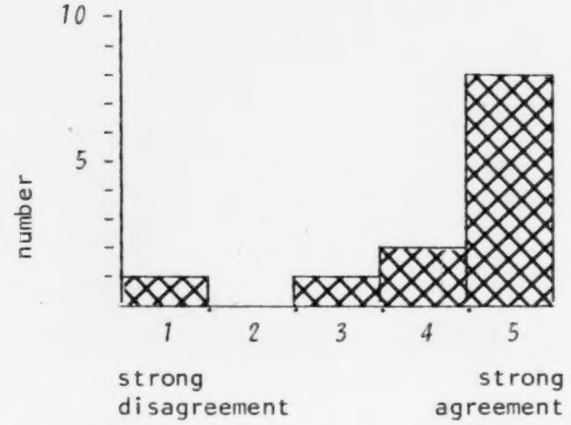


Fig. 9. Student responses on a 1 to 5 agreement scale to the statement, "I usually tried to be accurate and honest on the SRF."

The development of the SRF is still in only the preliminary phase. Revisions in its format, content, and use are made at the end of every academic term based on the data produced, and all data are now continuously computer processed. The demonstration of the validity of the SRF as a measuring tool is of paramount importance prior to further use, and the task is tactically difficult. However, such an empirical program is now in progress.

The remainder of this paper is concerned with the potential uses of the SRF. The data included are presented only as an example of the kind of information which might be available from its use and is not intended to be important in and of itself.

Research

The SRF had considerable utility and importance both as a research and as a teaching tool. There are literally hundreds of relationships which might be empirically examined in a single subject or group designs with study behaviors being either independent or dependent variables. One area of research with promising potential is the investigation of the effects on academic performance of varying the kinds and quantities of different study behaviors. An equally important approach would entail observation of the effects of different instructional procedures on study behaviors. The eventual goal of these and other research programs should be to empirically define the most effective and efficient study techniques for different subject matters and different teaching methods. This is in addition to the strictly heuristic use of the SRF in research to assist in examining other aspects of teaching.

Beyond its value as a research tool, the SRF offers an important means of improving our efforts at evaluation: evlauation of teachers themselves, evaluation of curricula, and evaluation of teaching methods.

Teacher Evaluation

The assessment of a teacher and his methods is one of the more complex tasks facing higher education. However, even more difficult and sensitive is the measurement of the effects on students of the teacher himself independent of the general techniques he or she uses--that is, the effects of a teacher's actions and personal interactions regardless of whether or not in general he is giving lectures and essay exams or individualized instruction and fill-in quizzes.

An example of the use of the SRF in this kind of evaluation may be seen by comparing the data from three student teachers using the same methods in the same course. This project was conducted in the Teaching Center at Georgia State University, Winter Quarter, 1971, on a junior-level course in abnormal psychology. The course was conducted in a manner similar to that described in Johnston and Pennypacker (1971), and Johnston and O'Neill (1972).

Very briefly summarized, each student in the course was assigned to another undergraduate (here called a Student-Teacher) who had previously completed the course with exceptional performance. After the student had studied a unit of text material, he signed up on his Student-Teacher's appointment sheet for a Performance Session at a particular time. Upon arriving he was given a confidential SRF (the original version) to complete. The student and his Student-Teacher then sat down in a booth, and the student was handed a variable number of fill-in questions typed on 3" x 5" cards. These 35 to 55 or so items were randomly drawn from a pool of 100 or more for each unit of subject matter in the course. After reading and answering aloud the stack of items or until a set period of time (often 10 minutes) had elapsed, the student and his Student-Teacher then engaged in an important required period of discussion of errors, text material, class material, etc. The student's rates of correctly and incorrectly reading and answering the fill-in items were computed and plotted on each student's own graph of his course performance. The student also wrote those rates on his SRF and deposited it in a locked box. If the student's performance unit met the course criteria he would begin working on the next unit. If not, he would come back to perform again (and again, if necessary) until he attained the criteria. Students who met or surpassed the criteria on every unit received an A in the course.

Table 1 presents the mean performance data and some SRF data for three different Student-Teachers who were using the same procedure in this course. Three Student-Teachers were chosen whose students' mean rates of correct and incorrect responding are highly similar so as to emphasize the differences in study tactics among the three groups. Given

TABLE 1

MEAN PERFORMANCE AND SRF DATA FOR THREE STUDENT-TEACHERS USING THE SAME TEACHING METHODS.

	Student-Teachers	A	B	C
	RATE CORRECT	3.6	3.5	3.3
	RATE INCORRECT	.8	.9	.9
	NO. PS/UNIT/STUDENT	2.3	2.6	3.1
	HOURS OF STUDY/PS	4.1	3.1	4.5
0 - 5 Ratings	READING FIRST TIME	2.3	2.5	2.4
	REREADING	2.5	1.3	2.7
	UNDERLINING	3.0	1.8	1.7
	OUTLINING	1.3	.3	.5
	SUMMARIZING	2.0	.3	.4
	TERMS, NAMES, DEFINITIONS	2.9	.6	.6
	SELF-QUIZZING-SILENT-TEXT	2.5	1.7	1.1
	SELF-QUIZZING-SILENT-WRITTEN AIDS	1.8	.7	.5

identical teaching methods and resulting performance, the data show some unanticipated and interesting differences in study behaviors among the three sets of students. There is a wide range of the number of Performance Sessions (PS) per unit per student between Student-Teachers A and C, and an even wider range (almost an hour and a half) of study time per Performance Session between Student-Teachers B and C. The reported proportion of time spent reading the material for the first time was quite constant across the three groups, but the students of Student-Teacher B spent only half as much time rereading as a study tactic compared to the other students. There are sizable differences across the three sets of students on the time spent in silently quizzing themselves on the text, and these ratings are inversely related to the number of performance sessions per unit per student, suggesting that this is an effective and efficient study behavior. Finally, the average ratings of the students of Student-Teacher A were generally higher than those of the other two groups.

Differences similar to these have frequently been observed among the other Student-Teachers of this and other courses. Although some differences could be in the students' academic backgrounds, an important question concerns what in the behavior of the Student-Teacher could produce such differences in the patterns of study of their students? It would seem that we should examine more closely the ways in which Student-Teachers, Proctors, Managers, etc., can and do influence the study tactics and performance of students. For instance, does *how* errors and text material are discussed make a difference? Does Student-Teacher *enthusiasm* make a difference? The SRF should be most useful in answering these and other questions.

Curriculum Evaluation

In addition to teacher evaluation, the SRF has considerable value as a tool to assist in curriculum evaluation. It has long been clear to teachers that the nature of the subject matter itself has measurable effects on both the student's academic performance and his study behavior.

An example of how the SRF can be used in this fashion may be seen in a comparison of a junior-level course in experimental psychology with the previously described abnormal psychology course. The experimental psychology course was conducted at Georgia State University using the same teaching methods as described above for the abnormal psychology course.

Table II shows the mean difference in performance and in some study behaviors for the two courses. Compared to the abnormal course, data from the experimental course show markedly better performance, almost a whole Performance Session less per unit on the average, and more than a half hour less study per Performance Session for each student. This resulted from the fact that students in the experimental course spent proportionately more time reading the material for the first time, rereading the material later, underlining and marking in the text, and writing down terms, names, and definitions.

While it is possible that some differences might have resulted because the two courses could have attracted students having consistently different study skills, similar differences have been regularly observed among courses of other subject matter and among different units in the same course. The example is intended to show how the SRF might be used to answer questions such as the following:

- ★*How do you empirically define the difficulty of material?*
- ★*Do two different texts for the same material produce different study patterns and performance?*
- ★*Should a teacher change some part of a course curricula?*
- ★*Are all units of equal difficulty?*
- ★*Do you need to emphasize certain study behaviors to get better performance with different kinds of material?*

TABLE II

MEAN PERFORMANCE AND SRF DATA FOR TWO DIFFERENT COURSES USING THE SAME TEACHING METHODS.

	Course	Abnormal	Experimental
	RATE CORRECT	3.2	3.8
	RATE INCORRECT	1.0	.8
	NO. PS/UNIT/STUDENT	2.9	2.1
	HOURS OF STUDY/PS	3.4	2.8
0 - 5 Ratings	READING FIRST TIME	2.4	3.1
	REREADING	2.4	3.0
	UNDERLINING	2.1	3.2
	OUTLINING	.8	.9
	SUMMARIZING	.9	1.0
	TERMS, NAMES, DEFINITIONS	1.5	1.9
	SELF-QUIZZING-SILENT-TEXT	1.8	1.7

Method Evaluation

One of the most valuable uses of the SRF is as a means of evaluating variations in teaching methods. An excellent example of such usage is in the SRF data collected as part of the Johnston and O'Neill (1972), study. This project was conducted with another section of the previously mentioned course in abnormal psychology using the same teaching procedures which have already been described. In this case, however, the independent variable manipulated was the minimum performance criteria for a course grade of A. There were five experimental conditions, two of which will be discussed here.

In Experiment I the students had no criteria to work toward; they were simply told that they would be graded on the curve along with others in their group. They were frequently urged to do their best. In Experiment II the criteria followed a *High*, *Medium*, *Low*, *High* sequence throughout the quarter, and the students were told that if they met the criteria on each unit (whatever it was) that they would get a course grade of *A*.

Table III shows mean performance and SRF data for both sections. In Experiment II it can clearly be seen that the quality of performance closely follows changes in minimum criteria. Of primary interest here, however, are the changes that occur in the students' study tactics as the criteria requirements vary. Almost all of the study behaviors show consistent changes related to the criteria shifts. Thus, the changes in performance result from not only changes in the quantity but the quality of study as well.

TABLE III

MEAN PERFORMANCE AND SRF DATA FOR EXPERIMENT I AND EXPERIMENT II

		EXPERIMENT II				EXPERIMENT I
	CRITERIA	Ⓗ	Ⓜ	Ⓛ	Ⓗ	NO CRITERIA
	RATE CORRECT	3.8	3.2	2.7	3.2	2.6
	RATE INCORRECT	1.0	1.2	1.2	1.0	1.5
	HOURS OF STUDY/ PERF. SESSION	3.0	2.4	2.2	2.8	3.5
0 - 5 Ratings	READING FIRST TIME	2.8	2.9	3.2	2.8	2.9
	REREADING (text)	2.7	2.5	2.4	2.8	2.4
	REREADING (written aids)	2.1	1.3	1.5	1.7	1.5
	UNDERLINING	2.5	2.8	2.9	2.6	2.9
	OUTLINING	1.1	.5	.8	1.3	1.3
	SUMMARIZING	1.3	.6	.8	1.5	1.6
	TERMS, NAMES, DEFINITIONS	1.2	1.3	1.2	1.5	2.5
	SELF-QUIZZING (silent, text)	1.7	1.2	1.2	1.3	2.8
	SELF-QUIZZING (silent, written aids)	1.3	.9	.8	1.2	1.6

Even more intriguing are the data from Experiment I. The averaged performance of these students was poorer than even that of the *Low* criteria phase in Experiment II; however, they studied considerably more than the other students under any criteria level, especially the *Low* criteria. Furthermore, they spent proportionately more time outlining, summarizing, writing down terms, and quizzing themselves than the students did under the *High* criteria in Experiment II. Their good study tactics in conjunction with their poor performance suggest the possibility that minimum performance criteria may have some effects on academic performance in ways other than through their effects on study habits.

Therapeutic Teaching

Probably the most exciting use of the SRF is for the strictly practical use of providing a precise means of engaging in therapeutic teaching for those students who are having performance problems in a course. Therapeutic teaching is the precise and individualized instruction of a student who is performing at less than his maximum by diagnosing the causes of the inadequate performance and bringing about the changes necessary to produce improved performance. While such features of these teaching techniques as continuous recording, direct recording, and individualized performances are at least prerequisites to therapeutic teaching, equally precise information regarding the study behaviors (which in great part produce the performance) is an absolute must.

A useful way of viewing the SRF data for a problem student throughout the quarter is to maintain a graphic profile of all performance and SRF data on a single easily read sheet. Figure 10 is an example of such a profile.

This is some of the data of a 38-year-old man who took the course in abnormal psychology in Winter Quarter, 1971. Beyond any problems he had with study tactics, he was in the midst of getting a divorce and was temporarily living with his parents. Although we were not making any organized and consistent therapeutic interventions at the time of this project, his profile shows quite clearly the kind of information the SRF can provide in this format. By the completion of Unit II, both the problem and at least some of the solutions were already evident. He was studying too little overall, particularly on second and subsequent attempts on a unit. He was spending too little time rereading the text (note the increase on Unit II and the improved performance) and no time reading supplementary materials. In addition, he was not using techniques of known effectiveness such as outlining, writing down terms and definitions, and quizzing himself on the material. These early conclusions were borne out in the last two or three units where marked changes occurred in most of these behaviors along with improved performance. In these later units one may observe increases in study time on second attempts, in rereading the text, in reading supplementary material, and in outlining the material. The availability of such a precise and complete

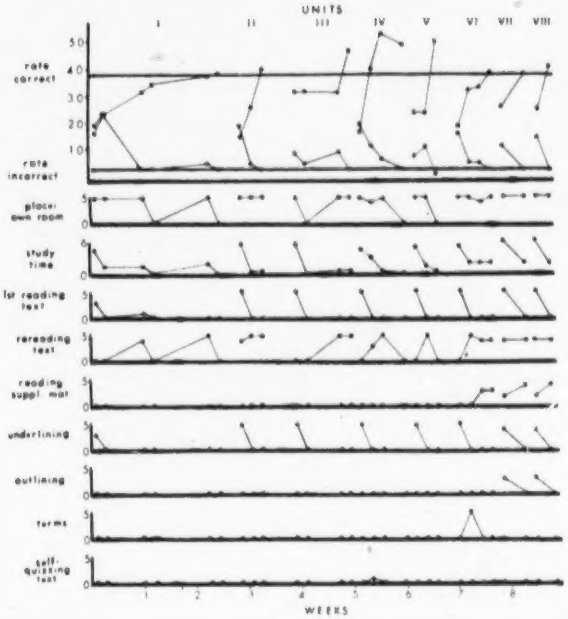


Fig. 10. Individual student performance and SRF profile.

record of academic performance and study behaviors for individual students showing poor performance in a course makes the early detection and amelioration of academic problems a relatively easy and precise task.

In summary, the SRF seems to hold profitable potential for providing extensive and accurate information about the study behaviors of the individual student for use in:

(1) any research project investigating teaching and performance;

(2) evaluating the effects of teachers themselves;

(3) evaluating the effects of curricula on students;

(4) evaluating teaching methods; and

(5) therapeutically instructing students who have performance problems.

REFERENCES

Johnston, J. M. and Pennypacker, H. S. A behavioral approach to college teaching. *American Psychologist*, 1971, 26 (3).

Johnston, J. M. and O'Neill, G. W. The analysis of performance criteria defining course grades as a determinant of college student academic performance. *Journal of Applied Behavior Analysis*, in press.

A MULTIPLE BASELINE ACHIEVEMENT TEST

l. keith miller
f. hal weaver

More and more university teachers are turning to the use of personalized instruction (Keller, 1968) and other behavior modification methods. These methods have in common that they specify the target behavior for the student; develop textual material utilizing the principles of fading, shaping, and stimulus discrimination; and program a variety of consequences capable of maintaining desired student behavior. The introduction of teaching methods based on behavioral research to the university classroom is certainly a gratifying development.

However, we have in front of us a major challenge: can we utilize our rigorous experimental methodology to evaluate--and hopefully, therefore, improve--the effectiveness of these teaching procedures? In spite of the great amount of work that many skilled behavior modifiers have put into this area, there exist surprisingly few experimental evaluations. To date, there are only statistical experiments in the literature comparing the final test scores of students enrolled in behavior modification sections with the test scores of students enrolled in traditional sections (Born et al., 1972; McMichael and Corey, 1969; Sheppard and MacDermot, 1970). These studies report higher test scores in the behavior modification sections than in the traditional sections.

The present paper reports the evaluation of a teacher-paced personalized instruction course by what might be called a *multiple baseline achievement test*. This procedure had three features. First, a comprehensive achievement test was administered to every student each of the 14 weeks of the semester. No feedback was given and no contingencies were placed on performance. Second, the test was constructed so that it had five subsections, each one covering one of the five major content units of the course. Each subsection was scored separately and these scores were used to define five separate behavioral baselines that might be affected by the experimental treatment. Third, the treatment for each baseline consisted of the implementation of the teaching package for the course unit corresponding to the content measured by that subsection of the achievement test.

Figure 1 shows the experimental design that this leads to. The dotted lines indicate measurements of achievement in subsections prior to the introduction of the teaching package--that is, the dotted line indicates baseline measures. For example, prior to the first week we measured

EXPERIMENTAL DESIGN

SUBSECTION	CONDITIONS
*RESEARCH METHODS	-----/TR PT PT PT PT
*REINFORCEMENT	---------/TR PT PT PT
*STIMULUS CONTROL	-------------/TR PT PT
*CONDITIONED REINFORCEMENT	-----------------/TR PT
*AVERSIVE CONTROL	---------------------/TR

KEY: ----- baseline
TR treatment
PT post-treatment

Fig. 1.

achievement in each of the five subsections. Next, we introduced the teaching package on the Research Methods part of the course. Thus, we have a measure of achievement after treatment on that subsection of the test (indicated by the TR after the slashed line); at the same time we continued to measure the achievement on the four untreated subsections. Thus, they remain in a baseline condition (as indicated by the continuation of the dotted lines). After about three weeks of Research Methods, we moved on to the unit covering Reinforcement. We then measured achievement on the Reinforcement subsection after treatment (as indicated by the slashed line and the letters TR). At the same time, we obtained a post-treatment measure on the Research Methods subsection as indicated by the letters PT; and we continued to gather baseline measures on the remaining three subsections that have not yet been taught. As these other units were taught we switched each subsection from a baseline to a treatment and then post-treatment condition. Thus, the teaching treatment is introduced sequentially to each of the subsections at approximately three week intervals. The result is a multiple baseline achievement test for the course.

This design was used to evaluate the achievement of 23 students enrolled in two sections of an introductory behavior modification course. The basic structure of the course might be described as a teacher paced personalized instruction course. The main contingency in the course was that students had to pass a daily quiz by answering 90% of the items correctly. If they did not pass the quiz they were required to take a make-up test until they did attain a 90% score. Their course grade was determined exclusively by the average grade on the initial quizzes.

The achievement test had five subsections, each containing ten fill-in-the-blank items. None of the items appeared on the quizzes. The test covered the basic principles that one might find in an introductory behavior modification course. The validity of the items was checked by administering the test to a sample of advanced graduate students. They scored an average of 89% with a range from 82% to 98%.

First, I am going to present the results of the achievement test in the traditional way. Figure 2 shows the before and after scores of the 23 students on the achievement test. Notice that on the average they scored 30% on the achievement test before taking the course. And, after they completed the course they scored an average of 85% correct. This is certainly evidence of a substantial increase in their ability to correctly answer questions about the basic principles of behavior modification.

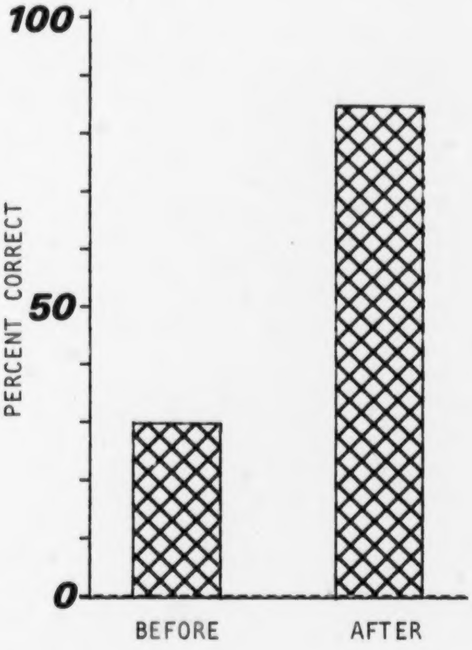


Fig. 2. Achievement score before and after introduction of the teaching package.

However, the question may still be asked whether their scores increased due to their exposure to the course. It might be argued that they had been exposed to this material in other courses in which they were enrolled. Or, it may be that they had read the course material, which was available to them throughout the semester, on their own. Or, it may even be that they became familiar with the test items and learned to answer them correctly from sheer familiarity. In short, before and after achievement tests do not demonstrate that any increase in achievement was *clearly* associated with the instructional methodology.

So, let me go back and present our achievement test results as a multiple baseline test. Figure 3 shows the results of the average score of each student on each of the subsections of the test plotted for each of the 14 presentations of the test. As you can see, the scores for each subsection range from a low of 22% correct for the Aversive Control subsection, to a high of 36% on the Reinforcement subsection at the time of the initial test. The average of all subsections on this before part of the test was 30%.

After the initial test, we assigned material on Methods of Research and started administering daily quizzes on the material. The average score during the next three weeks promptly jumped to 90% on the Methods subsection while remaining approximately stable on the other four subsections.

We next assigned materials on Reinforcement and started administering daily quizzes covering that material. The average score on the Reinforcement subsection jumped to over 80%. Meanwhile, the gain on the Methods subsection was retained; and the low scores on the other three sections held with some slight upward trend in the Conditioned Reinforcement subsection.

Next, we introduced the teaching package to the Stimulus Control subsection and the average score rose to over 80%. At the same time, the gains on the other sections held and again there was some slight upward drift evident in the Conditioned Reinforcement section.

Finally, we introduced the teaching package to the Conditioned Reinforcement section and next, to the Aversive Control section. Scores on both subsections increased, although the gain for the Aversive Control section was less dramatic than we have observed for the other sections. At the same time the gains registered on each of the baselines were retained.

And, as you already know, the final administration of the achievement test resulted in an average score of 85%, ranging from a low of 59% for the Aversive Control section to a high of 89% for the Methods of Research section.

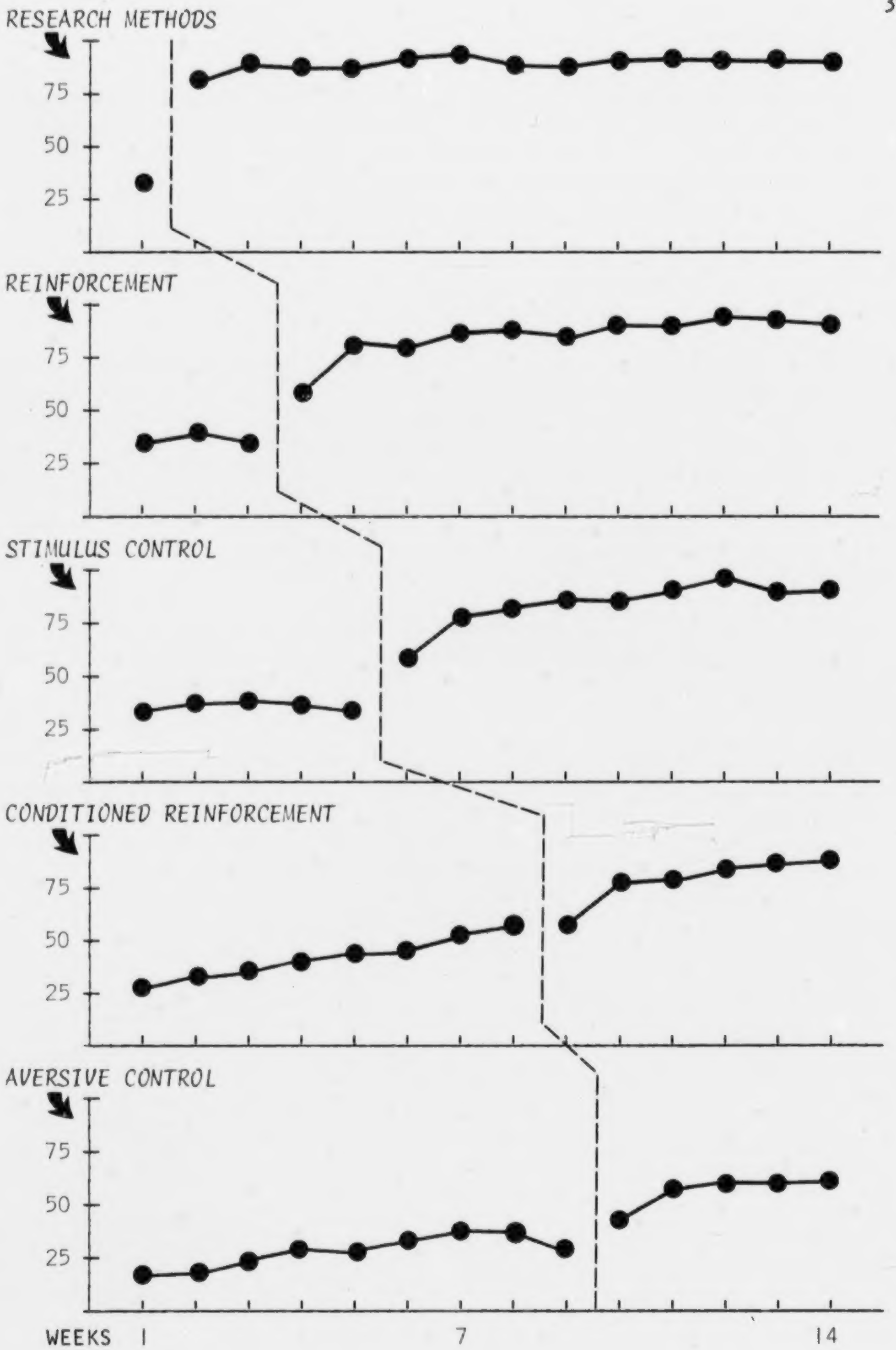


Fig. 3. Average scores on each subsection of the achievement test. Dotted line indicates introduction of teaching package for that subsection.

The multiple baseline achievement test clearly demonstrates that the increases in achievement for each subsection were closely correlated with the introduction of the daily quizzes and the grade consequences paired with them. It seems reasonable to conclude from such data that the gains in the achievement test were produced by the teaching package and not by study in other courses, self-directed reading of the material, or item familiarity. While this method did not permit us to break apart the teaching package and specify the relative importance of the quizzes, re-test requirement, or grade back-up, it clearly pointed to the effectiveness of the package.

What we have then is an addition to our measurement and design technology. The multiple baseline achievement test permits us to evaluate the effectiveness of the teaching package. It permits us to estimate the effect of other influences on student behavior such as item familiarity, other courses and free reading. It permits an assessment of the retention of material over a moderate period of time. And, it permits us to spot parts of the course that do not work as well as others--in this case the material on Aversive Control did not produce very high test results.

One further advantage of the design is that it gives us a technology for easily and continually evaluating the effectiveness of our teaching methodology without disrupting the course or losing time from productive teaching.

Let me close with the hope that this and other experimental methods such as we have seen today will be used to evaluate the different components of personalized instruction. Let us remember what happened to the programmed instruction movement as it floundered on the shoals of faith unaccompanied by hard nosed experimental evaluations.

REFERENCES

Born, D. G., Gledhill, S. M. and Davis, M. L. Examination performance in lecture-discussion and personalized instruction courses. *Journal of Applied Behavior Analysis*, 1972, 5, 33-43.

Keller, Fred S. Good-bye teacher... *Journal of Applied Behavior Analysis*, 1968, 1, 79-89.

McMichael, J. S. and Corey, J. R. Contingency management in an introductory psychology course produces better learning. *Journal of Applied Behavior Analysis*, 1969, 2, 79-83.

Sheppard, W. C. and MacDermot, H. G. Design and evaluation of a programmed course in introductory psychology. *Journal of Applied Behavior Analysis*, 1970, 3, 5-11.

◆

THE EFFECT OF TWO TEST - RETEST PROCEDURES ON THE CLASSROOM PERFORMANCE OF UNDERGRADUATE COLLEGE STUDENTS

darrel e. bostow
gerald j. blumenfeld

The present paper deals with the classroom behavior of undergraduate college students. Three studies are discussed. The first two provide evidence that deferring the credit students earn on course quizzes both increased the likelihood that students chose to take advantage of weekly remedial exams and increased the quality of their subsequent performance on a final exam. Details of the point system used to defer exam credit and a system for employing students in the scoring of their own exams are presented.

Recent experimental reports have presented several novel instructional techniques applicable to college courses (Keller, 1968; Ferster, 1968; Malott and Svinicki, 1969; McMichael and Corey, 1969; Sheppard and MacDermot, 1970; and Cooper and Greiner, 1971). Several of these studies have shown that more careful arrangement of reinforcement contingencies produces measurable differences in the performance of students when compared with more traditional college instructional techniques. The present research program was conducted to determine whether deferring credit for students who performed inadequately on the initial quiz of any given week in the course would both increase the likelihood that they would take a weekly remedial quiz and subsequently perform better on a comprehensive final exam in the course.

In the first experiment a large undergraduate course in introductory educational psychology provided 300 subjects. Content material in this and all subsequent courses discussed in this paper related primarily to elementary learning principles and their application in educational and home settings. Figure 1 presents the weekly format used in the courses. Many of the aspects of this method of instruction were developed by Dr. Jack Michael at Western Michigan University. As shown on the left side of the figure, all students received their first quiz of the week on Monday. Prior to each Monday, a set of specific behavioral objectives was handed out to the students describing in detail how they should prepare for the Monday quiz. A brief discussion period was carried on during the first few minutes of class to clarify problems in reading assignments. Immediately following the discussion period, all students took the quiz. As shown in the center of Figure 1, all

WEEKLY ACTIVITY SCHEDULE

Monday	Wednesday	Friday
Initial Quiz	Return Quizzes & Lecture-discussion	Remedial Quiz

Fig. 1. General weekly course format.

students returned to class on Wednesday and the graded quizzes were handed out for their inspection. During this time, the instructor clarified issues raised by the quiz and presented material which was not available to students in written form. Specific questions related to the quiz or general subject matter were first solicited from the students who had earned an *A* on the quiz. When questions from these students had been answered and new behavioral objectives for the following week were handed out, the "*A*" students were allowed to turn in their quizzes and leave. "*A*" students were excused from the Friday class. The remaining time in the Wednesday class was devoted to a remedial lecture - discussion aimed specifically at those students who remained in class because they had not earned an *A* on the first quiz of the week. The experimental programs discussed in the rest of this paper employed the present weekly activity schedule or a variation of it.

◆ EXPERIMENT I ◆

The first of the present research programs was designed to evaluate a system of deferring credit earned on the first quiz of the week when a student performed poorly. The rationale behind deferring credit being that it would encourage the student to restudy the assignment in order to perform better on the remedial quiz. For comparison purposes a large 300 student section of an undergraduate educational psychology class at Southern Illinois University was divided into two groups which experienced two testing conditions *alternately*. The experimental question which the first experiment set out to answer was: "If students perform poorly on a weekly quiz, does withholding course credit until they retake another quiz over the course material, make it more likely that they will return for the quiz than if credit was NOT deferred?" In other words, does delaying credit make students more likely to take the remedial quiz?

The weekly activity schedule in this first experiment was a slight variation from the general format outlined in Figure 1. In this schedule the course met four times each week. As shown on the left side of Figure 2, a lecture was given on each Monday. On Tuesday, the initial quiz of the week was given. For experimental comparison, half of the students in the course operated under one point condition, while the other half experienced a second point condition. We will call one of the point conditions the DEFERRED POINT SYSTEM. The left center portion of Figure 2 shows the system used for deferring credit. If a student operating under the deferred credit system earned a raw score of 16 to a maximum of 20 on the quiz, he was given 4 class points toward his final course grade. If he earned anywhere from 11 to 15 raw points on his quiz he received only 1 class point. If he earned 10 or less, he received no credit. As no more than 4 class points could be earned by a student in any given week, students coming to class on Wednesday and finding that they had earned maximum credit could not benefit from taking the remedial quiz on Friday. However, a student who earned a raw score of 15 or less on the initial quiz received so little credit for this performance that he needed to earn more points on the remedial quiz, or risk flunking the course. Class points in the course were cumulative; that is, the class points the student earned on each quiz were totaled at the end of the course in order to determine his letter grade. Thus, students performing poorly on an initial quiz were encouraged to take the remedial quiz or risk a bad letter grade at the end of the course.

As mentioned previously, half of the students experienced the DEFERRED POINT SYSTEM, while the other half of the students operated under a more conventional credit system. This system was termed the RAW SCORE SYSTEM because students, when operating under this condition, kept the raw score they earned on the initial quiz of the week. However, if a student were not satisfied with the raw score he earned on the initial quiz, he could elect to take the remedial quiz and replace his earlier raw

WEEKLY ACTIVITY SCHEDULE
Experiment #1

Monday	Tuesday	Wednesday	Friday
Lecture	Deferred Point System raw score / class pts. 16-20 / 4 11-15 / 1 10 or less / 0 Raw Score System raw score 1-20 Initial Quiz	Return Quizzes & Lecture-Discussion	Deferred Point System raw score / class pts. 17-20 / 3 14-16 / 1 13 or less / 0 Raw Score System raw score 1-20 Remedial Quiz

Fig. 2. Weekly course format for Experiment I.

score if his remedial score was higher. The basic difference between the DEFERRED POINT SYSTEM and the RAW POINT SYSTEM was that a student in the RAW POINT SYSTEM could retain a mediocre score on the initial quiz and elect *not* to take the remedial quiz. While, on the other hand, a student in the DEFERRED POINT SYSTEM lost a great deal of course credit if he performed poorly on an initial quiz and then neglected to take the remedial quiz.

The students serving as subjects for this experiment were randomly assigned to two groups. Throughout the experiment each group of students alternated between the DEFERRED POINT SYSTEM and the RAW POINT SYSTEM. In other words, each student, regardless of what group he was placed in, experienced both of the point systems. Figure 3 shows the experimental design. Notice by looking to the left side of the figure that Group A

EXPERIMENTAL DESIGN

$\frac{\text{ABAB}}{\text{BABA}}$

	Weeks 1 & 2	Weeks 3 & 4	Weeks 5 & 6	Weeks 7 & 8
DEFERRED POINT CONDITION	Group A	Group B	Group A	Group B
RAW POINT CONDITION	Group B	Group A	Group B	Group A

Fig. 3. Basic design for Experiment I.

experienced the DEFERRED POINT CONDITION during weeks 1 and 2, while at the same time Group B experienced the RAW POINT CONDITION. However, during weeks 3 and 4 each group was shifted to the opposite condition. During weeks 5 and 6, the groups were shifted again and so on. Shifting each group back and forth between the two point conditions allowed for both between-group and between-conditions comparisons, and circumvented student discontent resulting from staying in one experimental condition for the whole academic quarter. Figure 4 shows the system for equating credit earned during each experimental condition. The system was devised so that final letter grades could be given at the end of the course.

Total Points Accumulated* (CLASS POINTS (x5) plus raw score points)	Letter Grade
155-180	A
135-154	B
115-134	C
95-114	D
below 95	F

*180 was the maximum number of points possible and could be earned by a student taking nine weekly exams and earning the full 20 points on each quiz. Quiz performance data used in this study were taken from the first 8 weekly quizzes which allowed for four two-week experimental conditions.

Fig. 4. System for equating credit earned by individuals in Experiment I.

Results

Figure 5 shows results obtained during the academic quarter. In both weeks 1 and 2 a higher percentage of students in the group operating under the DEFERRED POINT SYSTEM returned to take the remedial exam. During the second two weeks, when the two groups were shifted to the opposite conditions, the same effect was observed. During weeks 3 and 4, both groups were shifted to the opposite point conditions. Thus, now a higher proportion of Group B students came back to take the remedial exam; these students were operating under the DEFERRED POINT SYSTEM. During the fifth and sixth weeks when both groups were shifted back to the original point system, a higher proportion of Group A elected to take the remedial quiz. This approximated the effect observed during the first two weeks of the course. Shifting the groups

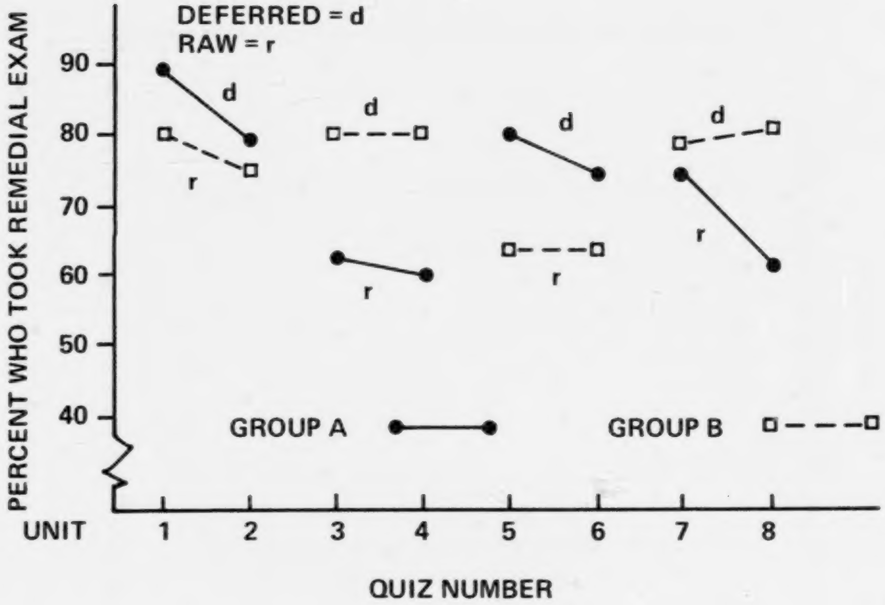


Fig. 5. Percent of students who elected to take the remedial quiz in Experiment I.

for the final time replicated the effect observed in the second two weeks. Results from the first experimental group evaluating the effect of deferring credit on the likelihood of students coming back for the remedial quiz, suggest that deferring credit *does* cause students to at least *take* the remedial quiz.

EXPERIMENT II

While results from Experiment I suggested that delaying course credit causes a larger precentage of students to take remedial quizzes, a second and more important question is whether forced remedial examination also affects quiz performance. To answer this question a second experimental program was conducted at the University of South Florida.

Figure 6 shows the weekly activity schedule in the second experimental program. This schedule was slightly altered from the one employed in Experiment I. In this second experimental program the introductory class in educational psychology met three days a week rather than four. Weekly quizzes in this course were expanded to 30 items incorporating both short answer essay and multiple choice quiz items. The criterion for *A* performance was raised to 90%. Also, a larger percentage of credit was deferred to the remedial quiz when a student performed poorly on the initial weekly quiz. The RAW POINT SYSTEM remained essentially the same as in the first experiment.

WEEKLY ACTIVITY SCHEDULE
Experiment #2

	Monday	Wednesday	Friday
Group A	Deferred Point System raw score — class pts. 27-30 — 10 21-26 — 2 18-20 — 0 below 18 — -1	Return Quizzes & Lecture-Discussion	Deferred Point System raw score — class pts. 27-30 — 8 21-26 — 6 18-20 — 4 below 18 — 0
Group B	Raw Point System raw score 1-30		Raw Point System raw score 1-30

Fig. 6. Weekly activity schedule in Experiment II.

The experimental design in this second program was also different. The class used as subjects in this study contained 54 students. The students were randomly placed into two groups prior to the beginning of the academic quarter. One group of students operated under the DEFERRED POINT SYSTEM throughout the entire academic quarter. The other group operated under the RAW POINT SYSTEM for the entire quarter. Both groups operated under their respective point systems and then at the end of the course both took the final exam at the same time. Quizzes were identical for both groups throughout the quarter. The final exam contained 60 multiple choice items with an equal number of items being selected from each weekly unit of content material.

Results

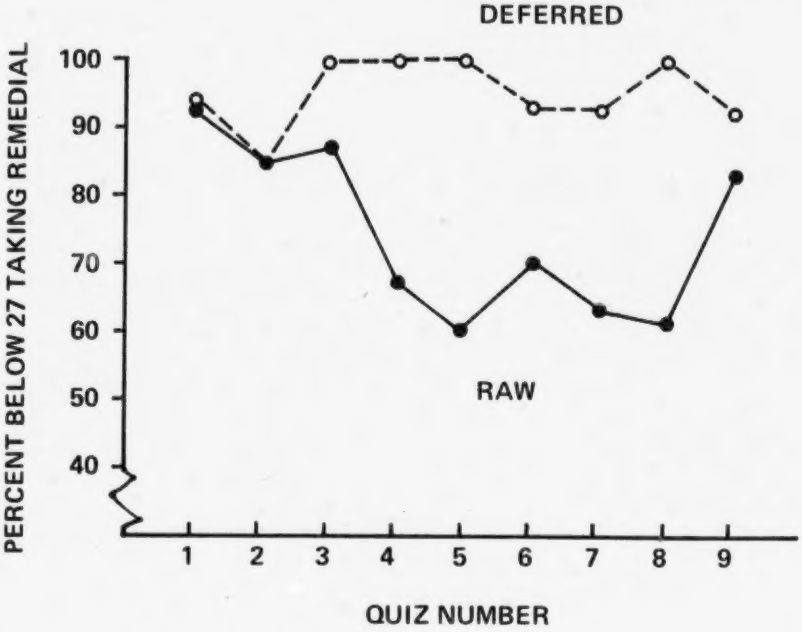
Data obtained in Experiment I showed that deferring course credit until the remedial exam increased the likelihood that a student would return for the remedial quiz. Figure 7 reveals that the same effect was observed in Experiment II. Data points in Figure 7 represent the percentage of students who came back to take the remedial quiz when they had earned less than an *A* on the initial quiz.

Figure 8 shows quiz averages for both groups on the initial quiz of each week. The average of test scores of students in the RAW POINT SYSTEM was higher than that achieved by the deferred credit group only in the first week of the academic quarter. On all subsequent initial quizzes in the course, the deferred group performed better. While more research seems necessary to clarify this effect, data in Figure 8 suggest that cumulative failure may have been occurring in the group of students who were less likely to make use of the remedial quiz each week.

As mentioned previously, all students received a 60 item final at the end of the course. Figure 9 presents the frequency distribution of raw scores on the final exam, for both experimental groups. The number of individuals in the upper distribution, the deferred credit group, was slightly smaller than the number in the second group. While students were randomly placed into the two experimental groups prior to the beginning of the academic quarter, several students whose names appeared on the preregistration class list failed to take the course causing slightly uneven experimental groups.

Figure 9 shows that the distribution of scores from the deferred credit group was slightly higher than those in the raw point group. This effect was found to be statistically significant when the Mann-Whitney U-Test was applied to the data ($Z = 3.20$, $p < .01$). Results from the second experimental program suggest that a system of deferring course credit when students performed poorly, thus making them more likely to make use of weekly remedial quizzes, served to increase the quality of their academic performance.

Fig. 7. Percent of students in Experiment II scoring below an *A* in each experimental group who elected to take the remedial quiz each week.



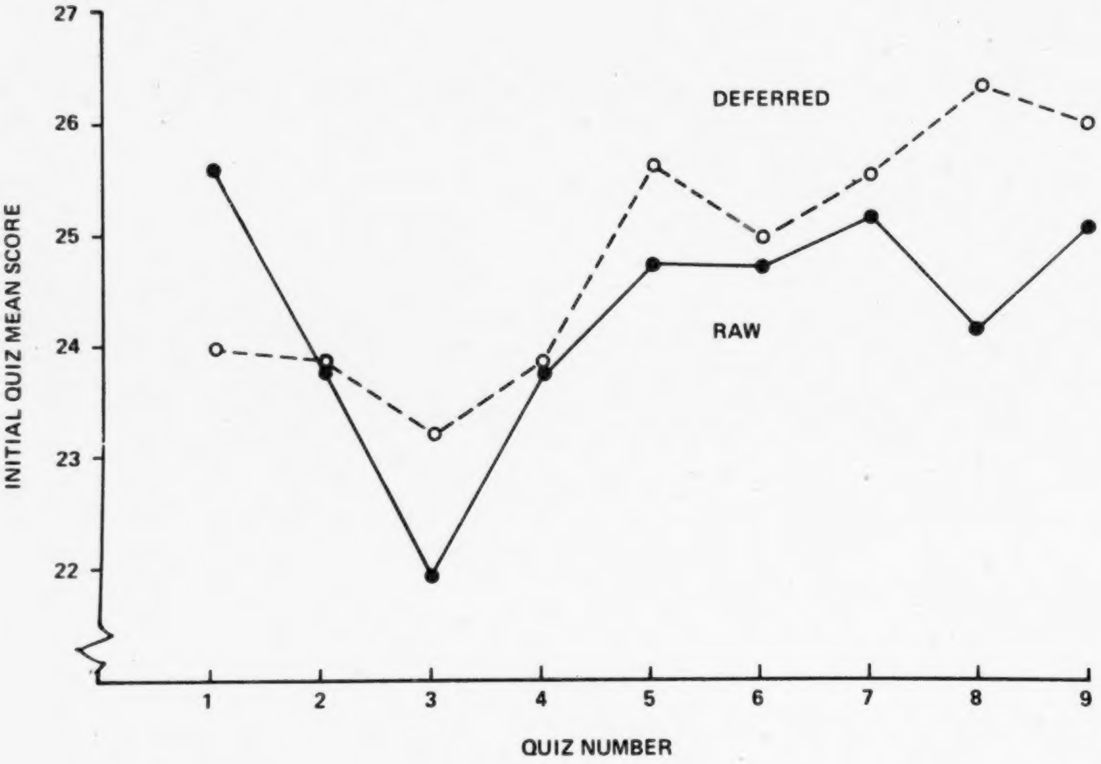


Fig. 8. Group quiz averages on the initial quiz of each week in Experiment II.

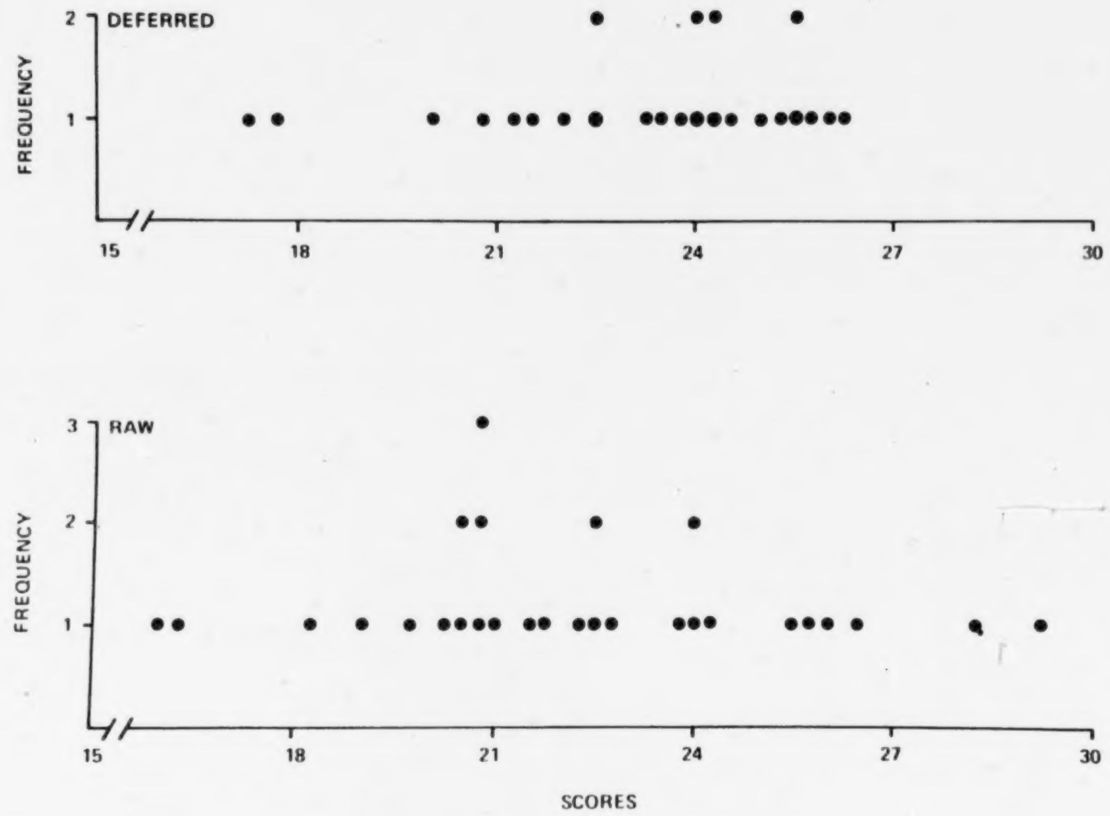


Fig. 9. Frequency distribution of raw scores on the final exam in Experiment II.

Quiz Grading

The third part of the present paper deals with the problem of getting the quizzes graded in courses similar to those in Experiment I and II. The following describes a method for utilizing students *from the class* to aid in the grading of quizzes. This quiz scoring system was designed to be incorporated into a DEFERRED CREDIT SYSTEM similar to that previously described. The following paragraph appeared in the course outline handed out to all students on the first day of class:

> As you might already have guessed, with many quizzes being given, there will be a lot of quizzes to grade. The following grading system has been used previously and was found to work quite well. I will grade the first two quizzes myself. Those students who earn a PASS on the first two initial quizzes will be placed into a pool out of which I will select a number of persons to help score quizzes. I will place the names of these students on a slip of paper which will be made available to students as they hand in their quiz papers on the third initial quiz of the course. If a student finds his name on this list, he can, if he wishes, help grade quizzes at the scoring session. If he elects to help grade, he automatically receives a PASS on the quiz he has taken. However, his quiz will be scored in the scoring session and I will note what each student would have gotten had he not received an automatic PASS. This grading system has a number of advantages. No student can reasonably anticipate that he will be able to score quizzes and receive an automatic PASS as he cannot find out whether he is eligible until after he has already taken the quiz. I have found that students who have received at least two "earned" PASSES in a row are usually quite able to correctly score other students' responses, especially when they have prepared to take the quiz themselves. I will announce a number of measures calculated to insure grader honesty and reliability when we begin the procedures. Each student becomes potentially eligible to be a grader when he has earned a pass on two initial quizzes in succession at any time during the course.

Three major steps were involved in the selection of graders. Step 1 was the tabulation of all students who earned *A's* on the initial quiz for two consecutive weeks. Step 2 involved randomly selecting 20 students (from a class of approximately 50 students) from those who had *A's* two weeks in a row. These names were placed on a list which was made available to each student as he turned in his next following initial quiz. If a student found his name on the potential grader list he was asked to place his initials by his name to show up for the grading session. Step 3 involved allowing only the first 10 students who elected to help grade to participate in the grading session. The reinforcer for helping grade exams was an *AUTOMATIC A* on the quiz the student had just taken, no matter how poor his true score on the quiz was. Placing 20 names on the list but employing only the first 10 who signed up insured that even if several students declined the opportunity to help grade, at least 10 students would show up for the session. Students were informed that once they initialed the list they would receive an automatic pass only if they attended the grading session. However, if they neglected to attend the session but had signed up to do so, they would receive a 0 on the quiz.

Exam grading was accomplished later on each test day. The instructor shuffled and distributed approximately 5 quizzes to each grader. All scoring marks were made on a separate cover sheet and graders placed their signatures on the scoring sheet for each of the quizzes they scored. Requiring grader signatures on each of the papers allowed for monitoring of grader reliability and insured that no grader was allowed to score his own quiz. An answer sheet was given to each grader at the beginning of the session. The group of graders scored each quiz item at the same time. Only after all graders had completed scoring a given item did the instructor signal for going on to the next item. The instructor circulated to aid in the interpretation of ambiguous quiz answers.

This method of selecting graders had several advantages. All graders were chosen from those students who had performed well on previous quizzes. Since the selection of students who were given the opportunity to help grade was done on a random basis, no student could predict whether he was going to get the opportunity for an automatic *A* and, therefore, was admonished to study carefully for each quiz. A second advantage was that the giving of bonus credit was *not* necessary in order to get students to help grade the quizzes. This system allowed a student to "buy insurance" in that when he took the opportunity to help grade quizzes, he was insured of obtaining an *A* on the quiz and thus avoided the possibility of having to attend a remedial quiz. This appeared to be a reinforcer for helping to grade quizzes.

Several steps were taken to evaluate the quality of student quiz grading. All student graders were given separate cover sheets for each quiz they graded, and were asked to do their scoring on this sheet only. These sheets were separated from the quizzes following the grading sessions and all quizzes were regraded by a single graduate assistant. In order to make a rough comparison of the similarity of scores obtained by student graders and the graduate assistant, the grand mean for student graders was compared with the grand mean obtained by the graduate assistant in each week of the course. Figure 10 shows the grand means obtained. Data in this figure suggest a high correspondence between student grader and graduate assistant scoring accuracy.

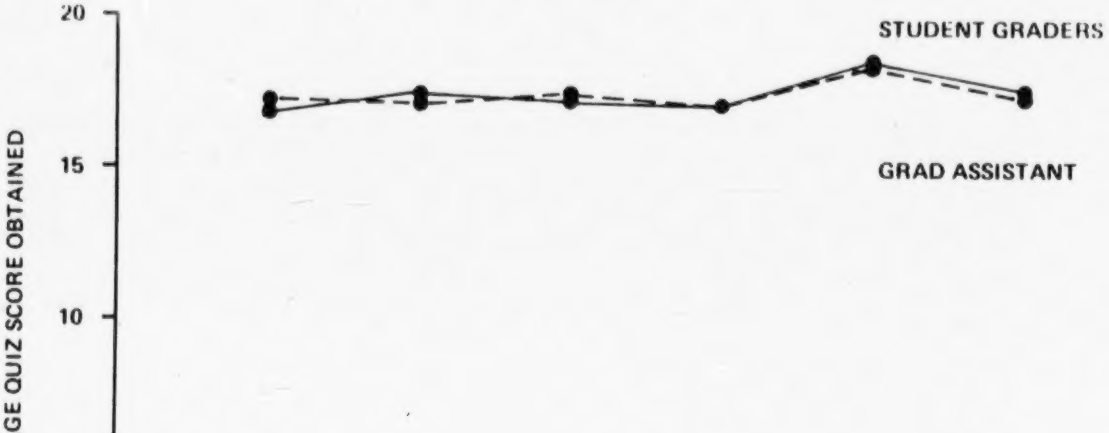


Fig. 10. Mean quiz score obtained by all student graders and mean scores obtained from the same quizzes by the graduate assistant.

All students who finished and turned in their quizzes before the total of 10 graders had been chosen were asked to inspect the list of potential graders. Approximately 3 out of 4 students who were eligible (by having two passes in a row) actually chose to participate in the grading sessions. Records were kept concerning how the students who declined to grade performed on the quizzes and these records were compared with the scores the graders would have earned had they not elected to grade. The mean score on the quizzes for which students received automatic passes was found to be 17.61 (17.00 or above was a pass; 20.00 was the maximum possible score during the academic quarter of this evaluation). However, the average score for students who had the opportunity, but declined to participate in the grading sessions, was 18.84. A t-test applied to these data indicated a statistically significant difference beyond the .01 level, suggesting that the students who elected to help grade exams and receive an automatic *A*, tended to be those who were more likely to benefit from the automatic *A*.

Summary

The preceding programs dealt with the classroom behavior of undergraduate students and have presented data suggesting that deferring quiz credit for poor student performance until the student takes a remedial quiz both increases the likelihood that he will take the remedial quiz and that he will subsequently perform better on the final exam. Especially when combined with the quiz scoring format described above, a deferred credit system appears to have significant educational advantages.

REFERENCES

Cooper, J. L. and Greiner, J. M. Contingency management in an introductory psychology course produces better retention. *Psychological Record*, 1971, 21, 391-400.

Ferster, C. B. Individualized instruction in a large introductory psychology course. *Psychological Record*, 1968, 18, 521-532.

Keller, Fred S. "Goodbye, teacher . . ." *Journal of Applied Behavior Analysis*, 1968, 1, 79-89.

Malott, R. W. and Svinicki, J. G. Contingency management in an introductory psychology course for one thousand students. *Psychological Record*, 1969, 19, 545-556.

McMichael, J. S. and Corey, J. R. Contingency management in an introductory psychology course produces better learning. *Journal of Applied Behavior Analysis*, 1969, 2, 79-83.

Sheppard, W. C. and MacDermot, H. G. Design and evaluation of a programmed course in introductory psychology. *Journal of Applied Behavior Analysis*, 1970, 3, 5-11.

All proceeds which derive from the sale of this volume, over the cost of printing, will be used to help support future Conferences on Behavior Analysis in Education.